

Mixed causal-noncausal models

Citation for published version (APA):

Telg, S. (2017). *Mixed causal-noncausal models: identification, estimation and inference*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20171206jt>

Document status and date:

Published: 01/01/2017

DOI:

[10.26481/dis.20171206jt](https://doi.org/10.26481/dis.20171206jt)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

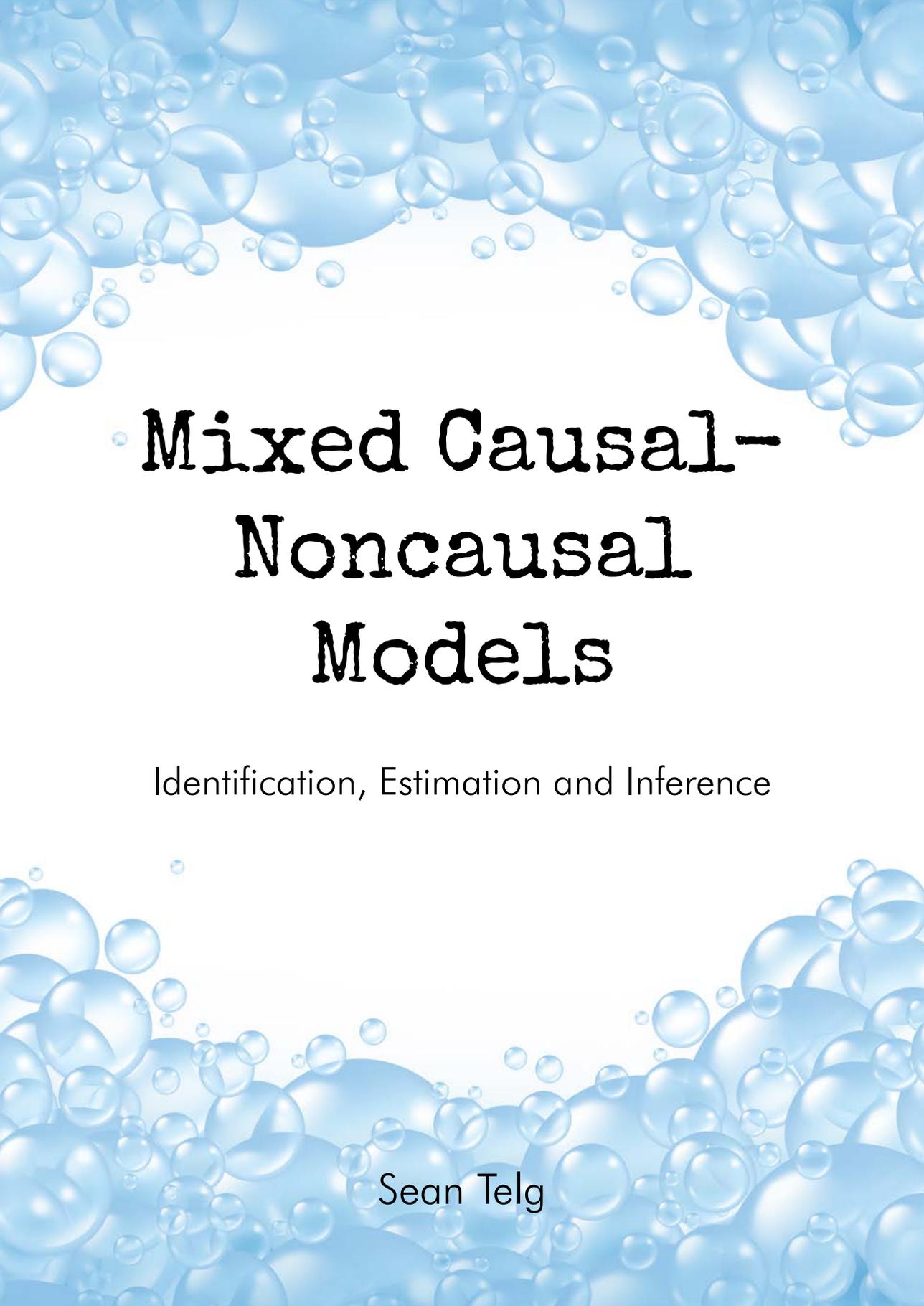
www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

The background of the slide is filled with numerous blue, translucent bubbles of various sizes, creating a bubbly, aquatic effect. The bubbles are more densely packed at the top and bottom edges, with a clear white space in the center where the text is located.

Mixed Causal- Noncausal Models

Identification, Estimation and Inference

Sean Telg

© Sean Telg, Maastricht 2017.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form, or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission in writing from the author.

This book was typeset by the author using L^AT_EX.

Published by Universitaire Pers Maastricht

ISBN: 978-94-6159-777-9

Printed in The Netherlands by Datawyse Maastricht



Mixed Causal-Noncausal Models

Identification, Estimation
and Inference

DISSERTATION

to obtain the degree of Doctor at
Maastricht University,
on the authority of the Rector Magnificus,
Prof. dr. Rianne M. Letschert,
in accordance with the decision of the Board of Deans,
to be defended in public
on Wednesday 6 December 2017 at 14.00 hours

by

Johannes Maria Andreas Telg

Supervisors

Prof. dr. A.W. Hecq

Prof. dr. J.R.Y.J. Urbain †

Co-Supervisor

Dr. L.M. Lieb

Assessment Committee

Prof. dr. F.C. Palm (Chair)

Prof. dr. H.P. Boswijk (University of Amsterdam)

Dr. S.T.M. Straetmans

Prof. dr. J.M. Zakoïan (CREST, Paris)

This research was financially supported by The Graduate School of Business and Economics (GSBE).

To my parents
who have always encouraged and supported me.

Acknowledgements

*“Feeling gratitude and not expressing it is like
wrapping a present and not giving it.”*

-William Arthur Ward (1921-1994)

In September 2009, I started my Econometrics adventure as a teenager who just finished high school, not completely convinced that he made the right study choice. Because of my indecisive nature, I only phrased the vague wish that “I wanted to do something with economics”. Combine this with the fact that I also liked mathematics and my choice was made. However, I had no idea what to expect. I remember opening one of the first handouts which stated “we assume you are familiar with this list of mathematical symbols” and I only recognized a handful of them. At that moment, I felt like I was going to face the biggest struggle and challenge ever in my life. I could have never anticipated that I would defend my PhD thesis almost exactly 8 years later. It was a long journey, but one that I definitely do not regret taking. The beginning of the bachelor was tough, but the challenges that I had to overcome and also successfully overcame intrigued me. It led to my decision of following a Research Master. Whereas I enjoyed doing research from the very start, I had some difficulties to really get started. I remember that I continuously had the same type of questions and doubts on my mind, like for example “is this novel (enough)?” or “maybe someone else is already doing this”. I started questioning whether pursuing a PhD would be a good idea. Looking back at it now, I am extremely happy I decided to go through with it after all. The past three years have been amazing for various reasons. I dare to say that I have improved considerably as both a researcher and a teacher. But most importantly, I believe I have grown a lot as a person. For this, I have many people to thank.

First of all, I would like to express my gratitude to my daily supervisors Alain Hecq and Lenard Lieb. It has been a pleasure working together with you over the past years. I have not only learnt a lot from both of you, but additionally, our meetings and conference trips have always been extremely enjoyable. Alain, thank you for always being there for me the numerous times I stood in front of your office door. Lenard, the meetings/coffee breaks/lunches/walks were always refreshing and fruitful, but above all a lot of fun. I could have never finished this PhD within the appointed three years without the both of you.

A warm word of thanks to my colleagues from the Quantitative Economics Department. Yolanda and Karin, the department is blessed with the two of you, as you are always happy to help and besides always in for a nice talk. Fellow (former) PhD candidates, you are too many to mention by name, but thank you all for the fun times we had. I would like to thank a few of you in particular. First off, Marina, my office mate: you have been with me on this journey from the

very beginning and it makes me sad to know it has come to an end. Alex, thank you for the several office talks, the various social events¹ and few workouts (way too few unfortunately) we did together. Étienne, thank you for an unfathomable amount of laughter. I am pretty sure we made our high school teachers proud. Hanno, my PhD-lifeline: thank you for sharing some of your infinite knowledge with me. Yuliya, especially the time in Paris has been wonderful and fortunately left us with a vast amount of inside jokes. Vincent, thank you for sharing a great amount of life experience and advice with me. Anna, I will miss your enthusiasm in everything you do. Shashwat, my football- and beerpong-buddy: it is always a blast. I could still mention so many of you, but let me just write: thank you all so much! This also goes for PhDs from other departments and the people from the Finance lab of CREST.

I want to thank my friends who possibly helped me in two different ways. Firstly, there is a big chance that I continued to talk about my PhD topic for way too long. My sincere apologies for this; sometimes it just helps to talk a lot regardless of whether anyone is actually listening. Secondly, we spent fun time together, which probably kept me from working on this thesis, which is also extremely important. Thank you to my friends from football, especially the indoor team that is cursed with always ending second place. It is always a pleasure to play a match with all of you and to drink a, usually well-deserved, beer afterwards.

A very special word of thanks and appreciation I would like to address to my family, in particular my parents. You have always supported me in every possible way and I know that I can always count on you. I hope you realize how important your role has been in me achieving this degree. I am truly blessed to have you close in my life. Of course, this also holds for my girlfriend Nina. Thank you for repeatedly listening to all the stories about mixed causal-noncausal models. You have been a great support for me in this journey.

Last, but definitely not least, I would like to mention one person in particular: Jean-Pierre Urbain. Sadly, you are not among us anymore while I defend this dissertation, but you are the person that convinced me to do both a Research Master and a PhD afterwards. You always believed more in me than I did myself. I am and will always be extremely grateful for that.

Sean Telg
Maastricht, October 2017

¹Our karaoke performance at Peter's Irish Pub being the main highlight.

Contents

“A person who won’t read has no advantage over one who can’t read.”

-Mark Twain (1835-1910)

Contents	xi
1 Introduction	1
1.1 Objective of the Thesis	3
1.2 Motivation	4
1.3 Applications	8
1.4 Outline of the Thesis	13
1.5 Notation	13
2 Identification of Mixed Causal-Noncausal Models in Finite Samples	15
2.1 Introduction	17
2.2 The Mixed Causal-Noncausal Model	19
2.3 Estimation	23
2.3.1 Student's t Maximum Likelihood Estimation	24
2.3.2 Least Absolute Deviation Estimation	26
2.4 Simulation Study	27
2.5 Empirical Applications	30
2.5.1 Empirical Illustration 1: Solar Panel Bubbles	30
2.5.2 Empirical Illustration 2: Realized Volatilities	36
2.6 Conclusion	39
2.A Weak White Noise	41
2.B Proof Covariance Matrix	42
2.C Equivalence of Information Sets	46
3 Do Seasonal Adjustments Induce Noncausal Dynamics in Inflation Rates?	49
3.1 Introduction	51
3.2 Mixed Causal-Noncausal Models	52
3.2.1 Model Representation	52
3.2.2 Estimation	53
3.3 Seasonal Adjustment Methods	54
3.3.1 The Linear X-11 Seasonal Filter	55
3.3.2 Properties of Seasonal Adjustment	56
3.3.3 Seasonal Adjustment for Mixed Processes	57
3.4 Simulation Study	58
3.4.1 Purely Causal and Noncausal Processes	58

3.4.2	Mixed Causal-Noncausal Processes	64
3.5	Seasonal Adjustment, Noncausality and Inflation Rates	67
3.5.1	Data	68
3.5.2	Results	70
3.5.3	Considerations	73
3.6	Conclusion	75
3.A	Autocovariances and Spectra	76
3.B	Graphs	78
4	Mixed Causal-Noncausal Autoregressions with Exogenous Re-	
	gressors	81
4.1	Introduction	83
4.2	The MARX Model	85
4.2.1	Simulation of MARX Processes	87
4.2.2	Identifiability	90
4.3	Maximum Likelihood Estimation	93
4.3.1	Asymptotic Properties of the AML Estimator	95
4.3.2	Computing the Covariance Matrix	97
4.4	Simulation Study	99
4.4.1	Performance MLE for MARX	99
4.4.2	Are MARX Models Identifiable Under Gaussianity?	100
4.4.3	Model Selection	102
4.5	Empirical Application	104
4.5.1	The Data	104
4.5.2	From Expectation Models to MARX	105
4.5.3	Identification of Pseudo ARDL models	108
4.5.4	Identification and Estimation of MARX	109
4.6	Conclusion	111
4.A	From Transfer Function Model to MARX	112
4.B	Approximate Likelihood Function	113
4.C	Proofs	114
5	Detecting Co-Movements in Noncausal Time Series	127
5.1	Introduction	129
5.2	Causal and Noncausal Models	131
5.2.1	Univariate Models	131
5.2.2	Noncausality and Common Features	132

5.2.3	Multivariate Models with Common Features	134
5.2.4	Co-Movements in Purely Causal or Noncausal Models	135
5.2.5	Co-Movements in Mixed Models	138
5.3	Test Statistics	139
5.3.1	LR Tests	139
5.3.2	GMM Tests	140
5.4	Monte Carlo Results	141
5.4.1	The Pseudo-Causal VAR(p) Order	141
5.4.2	Testing for Common Cyclical Features	142
5.5	Empirical Illustrations	147
5.5.1	Price Indicators	147
5.5.2	Commodity Prices Co-Movements	149
5.6	Conclusion	151
6	The MARX Package in R	153
6.1	Introduction	155
6.2	The MARX(r, s, q) Model	156
6.2.1	Model Specification	157
6.2.2	Simulation	158
6.2.3	Maximum Likelihood Estimation	160
6.2.4	Modelling Procedure	162
6.3	The R Package MARX	163
6.3.1	Structure	163
6.3.2	Simulation	165
6.3.3	Pseudo-Causal Models	166
6.3.4	MARX models	168
6.4	An Application on Commodity Prices	169
6.5	Conclusion	174
7	Conclusion	175
	Bibliography	181
	Nederlandse Samenvatting	193
	Valorization	199
	Curriculum Vitae	205

Chapter 1

Introduction

“Perplexity is the beginning of knowledge.”

-Khalil Gibran (1883-1931)

1.1 Objective of the Thesis

The analysis of time series data with a dynamic structure has been popular across many fields of research over the last decades. This has led to the development of a methodology which has become standard in the econom(etr)ic literature: the Box-Jenkins approach. This is a three-stage modelling approach that applies autoregressive moving average models to find the best fit of a time series to its own past values. The keyword for the existence of this thesis is the word *past* in the last sentence. It seems “natural” to only consider the dependence of an economic process on its own past, indicated as *causality*, especially from an intuitive point of view. However, in many other sciences (e.g., physics, signal processing and astronomy) it is common to consider models that are called *noncausal* or *acausal*. These are models in which the current value of a process depends on future realized values. Recent literature has shown that noncausal models, or more precisely a mixture of causal and noncausal models, can represent certain economic and financial processes better than causal models, due to the fact that they are able to take into account completely different features. Under certain assumptions, which are discussed extensively in the sequel of this chapter, it is sufficient to only consider causal models as they are indistinguishable from their noncausal and mixed causal-noncausal counterparts. However, if we relax these assumptions, the set of potential models to consider becomes much richer. Hence, the Box-Jenkins methodology can still be applied, but now acknowledges the possibility of a forward-looking component.

In this thesis, the mixed causal-noncausal autoregressive model (MAR) is studied and applied to various economic and financial processes. The purpose of this thesis is threefold. The first main objective is to make these models accessible to practitioners who wish to work with them. To that end, clear methodology on how to implement and interpret mixed causal-noncausal models is provided (Chapter 2) and the tools to implement proposed schemes are translated into a **matlab** routine and an **R** software package which are both freely available (Chapter 6). A second goal is to extend the existing model (Chapter 4) and to study the interesting features that come with this extension. These are considered from both an economic and statistical point of view. The third objective is to consider well-known econom(etr)ic topics and to consider them in the framework of mixed causal-noncausal models (Chapter 3, 5). The chapters in this thesis hope to contribute to this promising field within time series econometrics.

1.2 Motivation

The mixed causal-noncausal model is of interest to researchers for various reasons that are discussed in the sequel of this chapter. Before discussing its interesting features we start by explaining why the notion of noncausality has not been considered before and how it can be introduced in the Box-Jenkins methodology.

Box-Jenkins Method

In the Box-Jenkins approach, the variable of interest is first treated in such a way that it is stationary and seasonally adjusted. On the correctly transformed series, measures based on (partial) autocorrelation functions are used to decide on the number of autoregressive and moving average components necessary to capture the dynamics of the series adequately. As the proposed model is known at this stage, it has to be estimated using methods like maximum likelihood or nonlinear least squares. In a last step, misspecification tests are employed to assess whether the estimated model meets the specifications of a stationary univariate process. The Box-Jenkins approach only considers autoregressive moving average (ARMA) models which are dense within the class

$$y_t = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}, \tag{1.1}$$

which expresses a process y_t as a linear combination of current and past values of a white noise sequence ε_t . A justification for this can be found in the Wold Representation Theorem which dictates that every covariance-stationary sequence can be written as the sum of a stochastic part like in (1.1) and a deterministic part (like e.g. a sine wave or seasonal dummies). The theorem can be considered as an existence theorem and tells us that every weak stationary time series has a linear causal representation. This however need not be the best representation of a process. As an example, consider a process that is generated according to some nonlinear model of the form

$$y_t = f(y_{t-1}; \varepsilon_t), \tag{1.2}$$

with ε_t being a *strong* white noise sequence. If the process generated in (1.2) is covariance-stationary, then by applying the Wold Representation Theorem we know that this process also has a representation like in (1.1) where ε_t is only

required to be *weak* white noise. Hence, a linear model exists but it is not the only representation of the dynamic dependence of the series and might not be useful.

Noncausal Models

In a similar way, we can consider a noncausal autoregressive process. Take as an example, the noncausal AR(1):

$$y_t = \varphi_1 y_{t+1} + \varepsilon_t, \text{ with } |\varphi_1| < 1, \quad (1.3)$$

with ε_t strong white noise. The stationary solution of this model is a linear combination of current and future values of the white noise sequence ε_t given by

$$y_t = \sum_{j=0}^{\infty} \beta_j \varepsilon_{t+j}, \quad (1.4)$$

where $\beta_j = \varphi_1^j$. Note that the notions of “causality” and “noncausality” are defined with respect to the white noise process ε_t . That is, if a process y_t only depends on past and current values of ε_t like in (1.1) it is called causal, while it is called “noncausal” when it has a stationary representation in the form of (1.4). A process that depends on past, current and future values of the white noise sequence is called mixed causal-noncausal.¹

It can be deduced that in the case of (1.3) y_t is covariance-stationary, which means that the process can be represented as a causal process by the Wold Representation Theorem. Following Gouriéroux and Zakoïan (2016), if the process in (1.3) is modelled as a causal AR(1), which has a stationary solution given by (1.1), the error term is solely uncorrelated and not necessarily independent. In fact, Breidt and Davis (1992) show that the error term can only be independent in case of a Gaussian process. The Gaussian distribution is fully determined by second-order properties and due to the symmetry of autocovariance functions and spectral densities it is impossible to distinguish between causal and noncausal models. In terms of our example, if the error term in (1.3) is normally distributed, one does not lose any information by considering a causal AR(1). These models are completely equivalent and this gives a justification for only considering mod-

¹The terms “causal” and “noncausal” are not the best way to characterize these processes as they are not in line with the conventional definition of causality. However, as these expressions are widely used in the literature, we adhere to them in this thesis. Alternatively, one can speak of backward- and forward-looking processes.

els that are nested within (1.1). However, for non-Gaussian distributions, this is not the case: one can statistically discriminate between causal and noncausal specifications. The reason for this is that these distributions are not completely determined by second-order properties and also contain valuable information in other moments.

Estimation

The last observation explains why one should not use estimation methods like ordinary least squares for mixed causal-noncausal models. Because even if ε_t in (1.3) is non-Gaussian, the autocovariance function of the correct specification and a causal AR(1) are exactly the same. Since the estimation is completely based on this symmetric measure, the autoregressive coefficient in the causal AR(1) will be the same as the one in the noncausal AR(1). However, as discussed before, the noncausal AR(1) in (1.3) has a unique stationary solution of the form (1.4). If we express (1.3) in terms of a solution of the form (1.1) it can easily be seen that $|\alpha_j| > 1$, which means that the series y_t diverges instead of converges. This leads to the conclusion that any second-order estimator always computes the stationary solution. In terms of our example, if $\varphi_1 = 0.7$ in our noncausal AR(1) model in (1.3), the second-order estimation of a causal AR(1) leads to an estimated autoregressive parameter with a value close to 0.7. The estimator “falsely” suggests that the causal AR(1) has a stationary solution for this process, while this is clearly not the case.² For this reason, the non-Gaussian maximum likelihood estimator (Breidt et al., 1991), the least absolute deviation estimator (Wu and Davis, 2010) and the nonlinear covariance estimator (Gouriéroux and Jasiak, 2017b) are proposed to estimate mixed causal-noncausal models.

Mixed Causal-Noncausal Models

We have seen that there is no reason to only consider a one-sided causal filter as in (1.1) from a statistical point of view. To that end, for example Rosenblatt (2000) and Brockwell and Davis (1991, 2002) propose to rather consider linear processes of the form

$$y_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}. \tag{1.5}$$

²Of course it is highly advised (in any case) to perform diagnostic tests on the residuals after estimation to assess the adequacy of the model.

In summary, the Box-Jenkins approach is changed in the following ways:

- (i) y_t is allowed to have a two-sided moving average representation like in (1.5);
- (ii) The error term ε_t is strong white noise rather than weak white noise;
- (iii) The error term ε_t is non-Gaussian and possibly leptokurtic.

This suggests the use of mixed causal-noncausal ARMA models, but in this thesis we only focus on autoregressive processes. More specifically, we focus on models of the form

$$\phi(L)\varphi(L^{-1})y_t = \varepsilon_t, \quad (1.6)$$

where $\phi(L)$ is a polynomial of order r , $\varphi(L^{-1})$ a polynomial of order s with L being the usual backshift operator creating lags and the inverse of the lag operator, L^{-1} , creating leads. In order to ensure that the model has a two-sided stationary solution, the roots of the two polynomials $\phi(L)$ and $\varphi(L^{-1})$ are assumed to lie outside of the unit circle.

Testing the Normality Hypothesis

A necessary condition for identifiability of mixed causal-noncausal models is a non-Gaussian error term. In order to assess how restrictive this condition is for macroeconomic data, we investigate this property for the large FRED-MD database described in McCracken and Ng (2016). The dataset consists of monthly observations on 126 variables which mimic the coverage of datasets already used in the literature, but they add three appealing features: (i) they are updated in real-time through the FRED database, (ii) they are publicly accessible, facilitating the replication of empirical work and (iii) data changes and revisions are directly incorporated by the data desk at the Federal Reserve Bank of St. Louis. The monthly time series are divided into 7 groups: “output and income”, “labor market”, “housing”, “consumption, orders and inventories”, “money and credit”, “interest and exchange rates” and “prices”. Almost all variables span the period March 1959 - December 2016, which accounts for 694 monthly observations. There are only around 10 time series that are observed over a shorter time period, but they always have at least 250 observations each.

Empirically, the null hypothesis of normality can be tested as follows. As soon as the data series is correctly transformed such that it is stationary and seasonally adjusted (transformations for this dataset can be found in McCracken and Ng, 2016), one identifies the ARMA model that captures the dynamics adequately.

That is, one applies the original Box-Jenkins methodology as described above. The residuals resulting from fitting the chosen ARMA model are checked for normality. This can be done using well-known tests like e.g. the Jarque-Bera and Kolmogorov-Smirnov test. When the null hypothesis of normality is rejected, it is justified to investigate the presence of a noncausal component.

We identified ARMA models for all series in the FRED-MD database and performed the Jarque-Bera test on the corresponding residuals. At a 5% significance level, normality cannot be rejected in only 2 out of 126 cases. This means that 124 series are available for further investigation. It also highlights that the luxury assumption of normality in the original Box-Jenkins methodology is often not appropriate for macroeconomic data. Thus, the adaption of the method to impose a non-Gaussian error term seems not restrictive from this point of view.

1.3 Applications

The adjustment introduced in the Box-Jenkins approach enlarge the set of potential dynamic models to be considered. In this section, we consider different ways in which the mixed causal-noncausal model can be useful for (applied) researchers.

Non-Stationarity

The interest for mixed causal-noncausal models originated from a non-stationarity perspective. Breidt et al. (1991) considered an $AR(p)$ model of the form

$$a(L)y_t = \epsilon_t, \tag{1.7}$$

and proposed a maximum likelihood estimation procedure for the case that this model has r well-behaved roots (outside the unit circle) and s ill-located roots (inside the unit circle) with $p = r + s$. Lanne and Saikkonen (2011a) recognize the fact that the s roots are well-behaved in reverse time and thus split the polynomial $a(L)$ in two parts and form the model in (1.6). Hence, an $AR(p)$ process that consists of explosive roots can thus be rewritten as a mixed causal-noncausal model with solely well-behaved roots. Following Brockwell and Davis (1991, 2002) we can characterize this property by considering an autoregressive model of order 1:

$$y_t = \phi_1 y_{t-1} + \epsilon_t, \text{ with } |\phi_1| > 1. \tag{1.8}$$

By repeated substitution, one can write this model in its moving average representation in the form (1.1). However, this is not a stationary solution to this model as there is no converge due to the condition $|\phi_1| > 1$. Instead, one could rewrite (1.8) in the following way by rearranging terms and increasing every subscript by one:

$$y_t = \frac{1}{\phi_1}(y_{t+1} - \varepsilon_{t+1}), \quad (1.9)$$

which, in turn, by repeated substitution leads to the following moving average representation:

$$y_t = -\sum_{j=1}^{\infty} \phi^j \varepsilon_{t+j},$$

which is nested within representation (1.4). In fact, acknowledging that (1.8) is actually a noncausal model in disguise, makes it an invalid regression to perform. Since y_t has a forward stationary solution, we know that y_{t-1} can be represented as a linear process involving the term ε_t and thus y_{t-1} is not independent from ε_t , which violates the assumption of exogeneity between regressor and error term. To summarize, noncausal models offer the opportunity to rewrite a process with explosive roots in direct time into a process in reverse time that has roots outside the unit circle.

Modelling Nonlinear Features

Another useful application of noncausal models is their ability to generate certain dynamic patterns that cannot be created using conventional causal autoregressions. More specifically, Gouriéroux and Zakoïan (2016) show that the noncausal AR(1) is able to mimic a speculative bubble when the error term follows a fat-tailed distribution. To explain this phenomenon, consider the following example. Assume we have a causal and noncausal AR(1) with (for the sake of simplicity) a deterministic error term that equals 1 at time $t = 20$ and 0 otherwise. The autoregressive parameter equals 0.8 and the starting value $y_0 = 0$ for the causal AR(1) and the terminal value $y_{T+1} = 0$ for the noncausal AR(1). Then we have in the causal case:

$$y_{18} = 0.8y_{17} + \varepsilon_{18} = 0.8 \cdot 0 + 0 = 0;$$

$$\begin{aligned}y_{19} &= 0.8y_{18} + \varepsilon_{19} = 0.8 \cdot 0 + 0 = 0; \\y_{20} &= 0.8y_{19} + \varepsilon_{20} = 0.8 \cdot 0 + 1 = 1; \\y_{21} &= 0.8y_{20} + \varepsilon_{21} = 0.8 \cdot 1 + 0 = 0.8; \\y_{22} &= 0.8y_{21} + \varepsilon_{22} = 0.8 \cdot 0.8 + 0 = 0.64, \\&\vdots\end{aligned}$$

that is, we see a major increase followed by a slow decay. In the noncausal case, the pattern is as follows (we reverse time order for convenience):

$$\begin{aligned}y_{22} &= 0.8y_{23} + \varepsilon_{22} = 0.8 \cdot 0 + 0 = 0; \\y_{21} &= 0.8y_{22} + \varepsilon_{21} = 0.8 \cdot 0 + 0 = 0; \\y_{20} &= 0.8y_{21} + \varepsilon_{20} = 0.8 \cdot 0 + 1 = 1; \\y_{19} &= 0.8y_{20} + \varepsilon_{19} = 0.8 \cdot 1 + 0 = 0.8; \\y_{18} &= 0.8y_{19} + \varepsilon_{18} = 0.8 \cdot 0.8 + 0 = 0.64, \\&\vdots\end{aligned}$$

which is exactly the same as for the causal model but then in reverse time: we see an increase until $t = 20$ and then an immediate drop to zero. The paths of both processes are depicted in Figure 1.1. It can be seen that the path of the noncausal AR(1) indeed mimics a speculative bubble, as it slowly builds up and suddenly collapses. The value of the autoregressive parameter determines how strong the exponential growth of the bubble is. Gouriéroux and Hencic (2014) use this insight to model the Bitcoin/USD exchange rate which contains a bubble that bursted in April 2013. In our example, we used a deterministic error term for the sake of simplicity. Note that the need for a fat-tailed error distribution stems from the fact that an extreme value has to be drawn (in our example at $t = 20$) in order to create the paths discussed before. By combining the causal and noncausal dynamics in the mixed causal-noncausal model, one can create even more interesting patterns (like e.g. asymmetric cycles) that previously could only be obtained using highly nonlinear and complex models.³

³See Figure 2.1 in Chapter 2 for simulated examples.

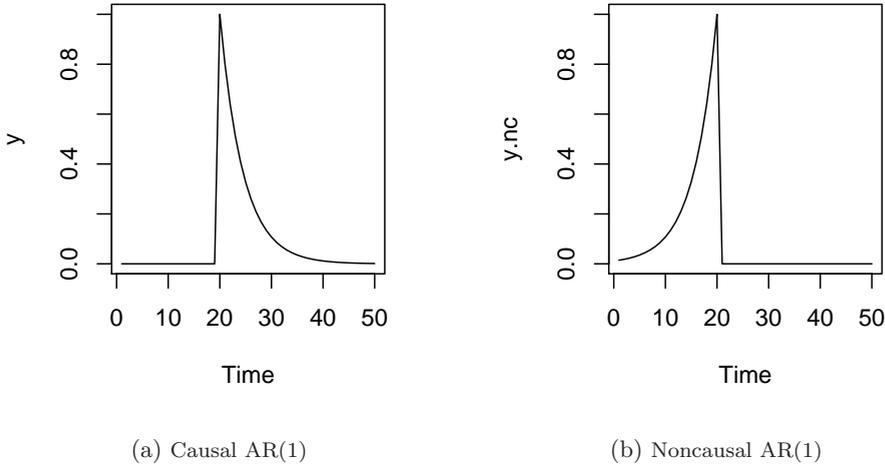


Figure 1.1: Baseline paths of the causal and noncausal AR(1)

Rational Expectation Models

Another interesting application of mixed causal-noncausal models is that they can empirically approach rational expectation models with nonfundamental solutions (Lanne and Saikkonen, 2011a). Following Hansen and Sargent (1991), the equilibrium of a typical dynamic rational expectation model is a covariance-stationary stochastic process which can always be represented in the form of an autoregressive model (of any finite order). This equilibrium can be fundamental or nonfundamental depending on the relationship between the dependent variable y_t and the white noise process ε_t . The definition of fundamentalness dictates that ε_t has to be contained in the linear space spanned by current and lagged values of y_t . That is, the current and lagged values of y_t have to be “fully revealing” about the ε_t in the moving average representation of y_t (Hansen and Sargent, 1991). From the conditions for fundamentalness stated in Alessi et al. (2011, p. 19), it follows directly that noncausal autoregressive models are nonfundamental by construction (due to their interpretation as an explosive causal autoregression). In other words, an economic process that depends on future shocks is nonfundamental. The detection of noncausality might be interpreted as the inadequacy of a causal autoregressive model to capture the process despite the fact that both models share the same

autocorrelation function and spectral density (Lanne and Saikkonen, 2011a). The information set of the econometrician who only observes current and past values of the process y_t is represented by a causal autoregressive model. A process that is best described by a noncausal or mixed causal-noncausal model can then be interpreted as the larger information set that agents have: despite the fact that these models have the same second-order properties as causal models with the same total autoregressive order, a model with a noncausal component is preferred since it additionally captures effects of omitted variables. A mixed causal-noncausal model allows for the explicit modelling of a process y_t on its future (expected) errors, which makes it a suitable alternative for the noninvertible moving average model which is often used in this strand of the literature. Gouriéroux et al. (2016) characterize how the noncausal stable AR(1) process is a stationary martingale or submartingale (depending on the value of the tail index) and thus constitutes to the set of stationary solutions with infinite variance for dynamic rational expectation models. Lanne et al. (2013) show how a specific example of a rational expectation model, the hybrid New Keynesian Phillips Curve, can be rewritten in terms of a mixed causal-noncausal model.

Forecasting

As explained earlier, the mixed causal-noncausal model can generate dynamic patterns that conventional causal autoregressive models cannot. Various papers (see e.g., Lanne and Saikkonen, 2011a; Gouriéroux and Hencic, 2014) show that a noncausal component indeed accommodates a better fit of several economic processes. These attractive features make them very suited for forecasting. Whereas it seems counterintuitive to forecast a process that depends on future values, Lanne et al. (2012a) and Gouriéroux and Jasiak (2016) propose methods to compute both point and density forecasts with mixed causal-noncausal models. These methods acknowledge two difficulties: (i) the one-step ahead predictor is generally nonlinear in the non-Gaussian noncausal case (Rosenblatt, 2000) and (ii) forecasting the noncausal component requires either simulation methods (Lanne et al., 2012a) or look-ahead type of estimators to compute predictive densities (Gouriéroux and Jasiak; 2016). The mixed causal-noncausal model has been found to provide better forecasts than conventional causal models (see e.g., Lanne et al., 2012a, 2012b; Lanne and Luoto, 2012). Despite the fact that forecasting is a very interesting field of application, it is not the focus of this thesis.

1.4 Outline of the Thesis

This thesis is structured as follows. In Chapter 2, the finite sample behavior of the maximum likelihood estimator (assuming a rescaled t -distribution) and the least absolute deviation estimator is studied. The limiting distribution of the causal and noncausal autoregressive parameters is specified for both estimators, which makes it possible to compute standard errors in closed form. The identification of mixed causal-noncausal models is illustrated in two empirical applications on the demand of solar panels and realized volatility measures for equity indexes. Chapter 3 studies the effect of seasonal adjustment on the identification of mixed causal-noncausal models. It is well-known that seasonal adjustment methods based on linear filters might induce spurious causal dynamics in conventional autoregressive processes and white noise series (Ghysels and Perron, 1993) and this chapter extends these results to mixed causal-noncausal processes. The results are confirmed in an empirical study on inflation rates of a large set of European countries. In Chapter 4, the mixed causal-noncausal autoregressive model is enriched with strictly exogenous regressors and the consistency of the non-Gaussian maximum likelihood estimator is proven. An empirical application on commodity prices in relation to the U.S. exchange rate and the industrial production index shows how this model can be seen as the solution of a rational expectation model. Chapter 5 considers the mixed causal-noncausal vector autoregressive process and shows that common features that cannot be detected in the causal representation might be revealed in the noncausal dynamics, due to the fact that these models are able to generate different features. Two empirical applications on price indicators from department stores and commodity prices reveal that additional relationships are discovered when one allows for causality and noncausality. In Chapter 6, the R package **MARX** is discussed. The software package offers functions to simulate, estimate and identify mixed causal-noncausal models based on procedures proposed in Chapter 2 and 4. Chapter 7 provides a short conclusion of this thesis.

1.5 Notation

Throughout this thesis, bold letters are used to indicate vectors and matrices. This convention is mostly chosen to make clear distinctions between polynomials, information sets and vectors. For example, $\phi(L)$ denotes the r th order lag polynomial $1 - \phi_1 L - \dots - \phi_r L^r$, while the bold letter $\boldsymbol{\phi}$ indicates the corresponding row

vector of coefficients $[\phi_1, \dots, \phi_r]'$. In case of multivariate polynomials, we keep this convention, despite the fact that they are matrix polynomials. For instance, $\Phi(L)$ indicates the r th order lag matrix polynomial $1 - \Phi_1 L - \dots - \Phi_r L^r$, while the separate matrices Φ_1, \dots, Φ_r in this polynomial are boldfaced. Vectors and matrices are always indicated by square brackets $[\]$, while information sets are denoted by round brackets $(\)$. That is, $\mathbf{y} = [y_1, \dots, y_T]'$ denotes a row vector of T observations, while $y = (y_1, \dots, y_T)$ denotes the information set of a process y_t . The notation y_t is used interchangeably to refer to a process $y_t = \{y_t\}_{t=1}^\infty$ or a single observation of this process at time t . The correct interpretation is assumed to be clear from the context.

Chapter 2

Identification of Mixed Causal-Noncausal Models in Finite Samples

By means of Monte Carlo simulations, we investigate the identification of mixed causal-noncausal models in finite samples for different values of the excess kurtosis of the error process. We compare the performance of the MLE, assuming a t -distribution, with that of the LAD estimator that we propose in this chapter. Similar to Davis et al. (1992) we find that for infinite variance autoregressive processes both the MLE and LAD estimator converge faster. We further specify the general asymptotic normality results obtained in Andrews et al. (2006) for the case of t -distributed and Laplacian distributed error terms. We first illustrate our analysis by estimating mixed causal-noncausal autoregressions to model the demand for solar panels in Belgium over the last decade. Then we look at the presence of potential noncausal components in daily realized volatility measures for 21 equity indexes. The presence of a noncausal component is confirmed in both empirical illustrations.¹

¹This chapter is based on the article Hecq A., Lieb L. and S. Telg (2016), Identification of Mixed Causal-Noncausal Models in Finite Samples, *Annals of Economics and Statistics*, 123/124, 307-331.

2.1 Introduction

Several time series textbooks (e.g., Brockwell and Davis, 1991, 2002) advocate the use of noncausal (or similarly, forward-looking) models. The reason for this is that these models offer the possibility to rewrite a process with explosive roots into a process in reverse time with roots outside the unit circle. Also in the applied econometric literature there is a growing interest in noncausal models. We see at least three reasons explaining this interest. First, several authors (see e.g., Lanne et al., 2012a, 2012b) have shown that noncausal models, which have the feature that they encompass information about the future path of stationary time series, might improve forecast performances obtained with purely causal autoregressive models. Most papers look at inflation in their evaluations, but also commodity prices (Lof and Nyberg, 2017) as well as other macroeconomic variables (Lanne et al., 2012b). Second, noncausal models are together with non-invertible models, special cases where shocks are non-fundamental (see Lippi and Reichlin, 1993a, 1993b; Alessi et al., 2011; Beaudry et al., 2015). This issue is crucial in macroeconometrics (Hansen and Sargent, 1991) and is interpreted as the evidence that the econometrician has a smaller information set at his disposal than the economic agents have. Third, and this will be the angle taken in this chapter, forward-looking autoregressive models with non-Gaussian fat-tailed disturbances are able to replicate features that previously could only be obtained by highly nonlinear and complex models. Gouriéroux and Zakoïan (2016) and Hencic and Gouriéroux (2014) have shown that a simple noncausal AR(1), i.e., $y_t = \rho y_{t+1} + \varepsilon_t$, with an *iid* α -stable distributed error term, is able to mimic processes similar to bubbles observed in economic and financial variables. Without entering into definition details (see e.g., Scherbina, 2013 or Stiglitz, 1990 for a survey), the term bubble roughly refers to as a rapid acceleration of prices (assets, real estate, commodity, etc.), over a natural level (often called the intrinsic value) given the fundamentals of the economy and discounted present values of expected returns. In general, this is driven by mechanisms such as speculations or anticipations. Because it is difficult to observe intrinsic values in real time, bubbles are often conclusively observed and identified in retrospect, when a sudden drop in prices appears. Such a drop is known as a crash or a bubble burst. However, standard causal autoregressive models were found to be incapable of capturing the behavior of series that increase and then suddenly burst. Nonlinear processes with a mixture of unit and explosive roots are often considered in the literature to model this bubble feature (see Phillips

et al., 2011 or the survey in Homm and Breitung, 2012). Noncausal models can provide a simple and convenient alternative representation. Moreover, they can be exploited to forecast turning points (crashes say) for different horizons and consequently noncausal models yield helpful risk measurements.

This chapter provides several elements that would facilitate the interpretation and the estimation of mixed-causal and noncausal models for practitioners. The few existing applied econometric papers² on purely noncausal and mixed causal-noncausal models are often not very specific on how to implement the estimation and the construction of inference (in a very practical, applied sense). Also the existing literature lacks a finite sample assessment of the (relative) performance of different approaches. We want to guide the practitioner on the following issues that we emphasize in this chapter: (i) We show that the fatter the tails of the distribution are, the faster the convergence of the estimator is achieved (and the more accurate the identification of the model is). This means that estimating causal and noncausal models for volatile financial time series will be, in theory, relatively easy; (ii) We compare the performance of a Student's t Maximum Likelihood Estimation (MLE) approach (advocated by Breidt et al., 1991 or Lanne and Saikkonen, 2011a) with an alternative Least Absolute Deviation (LAD) estimator and offer practitioners a straightforward procedure to compute standard errors of parameters which is not based on computing the Hessian using gradient based (numerical) procedures. We show by simulation that the LAD approach has appealing properties for odd distributions such as asymmetric ones or mixtures of distributions that are commonly encountered in practice; (iii) We also investigate how fast the causal and noncausal parameters (by MLE or LAD estimator) approximate their asymptotic normal distribution; (iv) As a first illustration, we provide the details of the necessary steps to implement those tools for the demand of solar panels in Belgium. While that application might be affected by the small sample that is used (about 75 observations), we provide a second application where we look at mixed causal-noncausal models on daily realized volatilities for 21 equity indexes measured over 2000-2014 and hence on more than 3000 observations; (v) We have developed a **matlab** and **R** toolbox that easily estimates and identifies those models.

The rest of this chapter is as follows. Section 2.2 recalls the elements needed to understand the literature on noncausal models. In Section 2.3 we elaborate

²For a comprehensive overview of the literature on this topic, visit Markku Lanne's (<http://blogs.helsinki.fi/lanne/>) and Joann Jasiak's (<http://dept.econ.yorku.ca/jasiakj/>) web-pages.

on estimation and how to construct inference on mixed causal-noncausal models. The results from the Monte Carlo simulations are collected in Section 2.4. Section 2.5 details the two applications. Section 2.6 concludes. An appendix collects additional material.

2.2 The Mixed Causal-Noncausal Model

The univariate mixed causal-noncausal autoregressive model, denoted $\text{MAR}(r, s)$ (see inter alia Lanne and Saikkonen, 2011a; Gouriéroux and Jasiak, 2016) is usually written as

$$(1 - \phi_1 L - \dots - \phi_r L^r)(1 - \varphi_1 L^{-1} - \dots - \varphi_s L^{-s})y_t = \varepsilon_t, \quad (2.1)$$

$$\phi(L)\varphi(L^{-1})y_t = \varepsilon_t, \quad (2.2)$$

with L being the backshift operator, i.e., $Ly_t = y_{t-1}$ gives lags and $L^{-1}y_t = y_{t+1}$ produces leads. When $\varphi_1 = \dots = \varphi_s = 0$, the process y_t is a purely causal autoregressive process, denoted $\text{MAR}(r, 0)$:

$$\phi(L)y_t = \varepsilon_t. \quad (2.3)$$

Model specification (2.3) can be seen as the standard backward-looking AR process, with y_t being regressed on y_{t-1} up to y_{t-r} . The process in (2.2) is a purely noncausal $\text{MAR}(0, s)$ model

$$\varphi(L^{-1})y_t = \varepsilon_t, \quad (2.4)$$

when $\phi_1 = \dots = \phi_r = 0$. Model specification (2.4) is the counterpart of (2.3), since it is a purely forward-looking AR process. That is, y_t does not depend on its past values, but rather on its future values y_{t+1} up to y_{t+s} . Models of the form (2.2) that contains both lags and leads of the dependent variable, are called mixed causal-noncausal models.

The roots of both the causal and noncausal polynomial are assumed to lie outside the unit circle, that is $\phi(z) = 0$ and $\varphi(z) = 0$ for $|z| > 1$ respectively. These conditions imply that the series y_t admits a two-sided moving average (MA)

representation

$$y_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \quad (2.5)$$

such that $\psi_j = 0$ for all $j < 0$ implies a purely causal process y_t (w.r.t. ε_t) and a purely noncausal model when $\psi_j = 0$ for all $j > 0$ (Lanne and Saikkonen, 2011a). Error terms ε_t are assumed *iid* (and not only weak white noise) non-Gaussian.

Following Gouriéroux and Jasiak (2016), we define the unobserved causal and noncausal components of the process y_t as follows:

$$u_t \equiv \phi(L)y_t \leftrightarrow \varphi(L^{-1})u_t = \varepsilon_t, \quad (2.6)$$

and

$$v_t \equiv \varphi(L^{-1})y_t \leftrightarrow \phi(L)v_t = \varepsilon_t. \quad (2.7)$$

The specification of these filtered values is very useful in simulating, estimating and forecasting mixed causal-noncausal processes.

The non-Gaussianity assumption ensures the identifiability of the causal and the noncausal part (Breidt et al., 1991). Most papers by Lanne, Saikkonen and coauthors use $t(\nu)$ -distributions with $\nu \geq 2$ as an alternative to the Gaussian distribution while Gouriéroux and coauthors rely on the α -stable distribution (in practice, mostly Cauchy or a mixture of Cauchy and Normal distributions). However, for most applications it emerges that the Cauchy distribution is too fat tailed and many series would have (if estimated by a Student's t MLE) a degree of freedom between 1.5 and 2.5.

Figure 2.1 characterizes how purely causal and purely noncausal AR processes of the same order can create a different structure when a non-Gaussian error distribution is assumed (for $T = 100$). The first two graphs (2.1a, 2.1b) show simulated paths³ for both a causal and noncausal AR(1) with an autoregressive parameter value of 0.9. In both cases, ε_t is assumed to be standard Cauchy, i.e., a $t(1)$ -distribution. The pattern of the peaks in these graphs reveal the main difference: (i) in the causal case, a jump is followed by an exponential decrease and

³MAR(r, s) models are simulated as follows. First generate a sequence of ε_t . Assume terminal values for u_t and construct the rest of the sequence according to $u_t = \varphi_1 u_{t+1} + \dots + \varphi_s u_{t+s} + \varepsilon_t$. Subsequently, y_t can be constructed, since $y_t = \phi_1 y_{t-1} + \dots + \phi_r y_{t-r} + u_t$ and only starting values for y_t are required. A burn-in period for u_t and y_t is recommended to remove the effects of the initial values.

consequently cannot mimic the typical bubble pattern, while (ii) in the noncausal case, one observes a process of exponential growth followed by a drop. The two remaining graphs (2.1c, 2.1d) show processes simulated from a model consisting of both a forward- and backward-looking component, namely the mixed causal-noncausal MAR(2,2) model. It can be seen how the combination of a causal and noncausal component can create asymmetric bubbles/cycles which encompass the features of both pure models discussed previously. This proves helpful for macroeconomic time series to describe asymmetric business cycles (Ramsey and Rothman, 1996), for instance when recession and recovery phases have different amplitudes. By means of choosing the sum of elements in the parameter vector ϕ (respectively φ) close to unity it is possible to increase the causal (respectively noncausal) polynomial as driving force in the process.

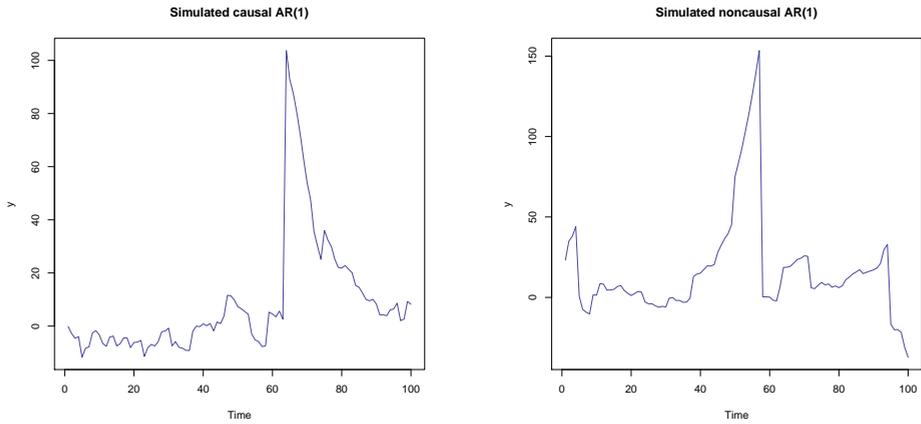
In the early literature, mixed causal-noncausal models were introduced as a solution to ill-located root problems. Following Breidt et al. (1991), an autoregressive model of order $p = r + s$, i.e.,

$$a(L)y_t = e_t, \tag{2.8}$$

with r roots outside the unit circle, s roots inside the unit circle and e_t an *iid* error process, can be rewritten as an MAR(r, s) model defined in (2.2) which has all roots strictly outside the unit circle. Hence, both models have a unique strictly stationary solution as a two-sided moving average representation given in (2.5). Appendix 2.A considers the particular case of a noncausal AR(1). It is however well-known that one can define an alternative model, say

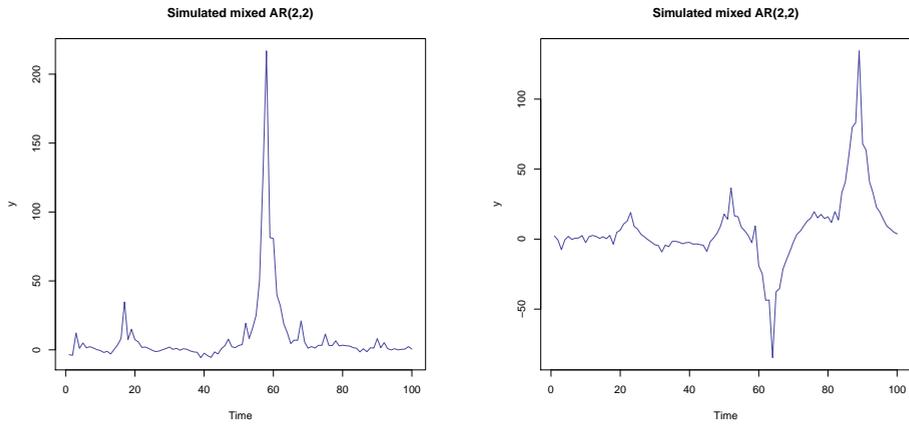
$$a^*(L)y_t = e_t^*, \tag{2.9}$$

which has all p roots outside the unit circle, while still maintaining the same autocovariance structure as the models in (2.2) and (2.8). Hence, any estimation method based solely on first and second order moments of the system is unable to distinguish among causal, noncausal and mixed causal-noncausal models. More specifically, if y_t is a noncausal AR(p) driven by an *iid* sequence of disturbances with mean zero and variance σ^2 , then y_t can also be expressed as a causal AR(p) driven by a newly created white noise sequence with mean zero and variance $\tilde{\sigma}^2$ (Brockwell and Davis, 1991). In the non-Gaussian case, the newly created white noise sequence is still uncorrelated, but not independent anymore (see Appendix



(a) Simulated causal MAR(1,0) process ($\phi_1 = 0.9$) with standard Cauchy errors

(b) Simulated noncausal MAR(0,1) process ($\varphi_1 = 0.9$) with standard Cauchy errors



(c) Simulated MAR(2,2) process ($\phi = [0.2, 0.3]'$, $\varphi = [0.2, 0.1]'$) with standard Cauchy errors

(d) Simulated MAR(2,2) process ($\phi = [0.3, 0.3]'$, $\varphi = [0.4, 0.4]'$) with standard Cauchy errors

Figure 2.1: Simulated processes from various MAR(r, s) specifications, $T = 100$

2.A for an illustration). We refer to model (2.9) as the pseudo-causal model,⁴ since even if y_t is purely noncausal or mixed causal-noncausal, its spectral density and thus autocovariance function cannot be distinguished from those of a causal autoregressive process.

⁴Gouriéroux and Zakoïan (2016) refer to the pseudo-causal model as the empirically weak causal linear model.

For this reason, Lanne and Saikkonen (2011a) propose to first estimate conventional causal autoregressive models by means of OLS or Gaussian MLE and to use usual information criteria to determine the lag order p . Subsequently, the combination (r, s) within $p = r + s$ that maximizes the log-likelihood is found to constitute the most appropriate mixed causal-noncausal model. Alternatively, Hencic and Gouriéroux (2014) propose to look at the empirical ACF which is consistent even in the infinite variance case to determine p . In both cases, diagnostic tests can be performed to investigate whether p lags are sufficient to capture the autocorrelation in the error process adequately.

2.3 Estimation

Let us assume that we estimate a purely causal autoregressive model and we reject both the Gaussianity and the *iid*-ness null hypothesis on the residuals. These test results can be interpreted as evidence that investigating a noncausal or mixed causal-noncausal specification is justified. In this section, we outline the estimation strategy of mixed causal-noncausal models as specified in Section 2, with the small modification that these models are augmented with an intercept α . If the error process ε_t possesses a density $f_\varepsilon(\varepsilon_t; \boldsymbol{\lambda})$, with $\boldsymbol{\lambda}$ being a vector collecting distributional parameters (e.g., scale parameters, degrees of freedom), the associated approximate likelihood function⁵ is given by

$$\prod_{t=r+1}^{T-s} f_\varepsilon(\alpha, \boldsymbol{\phi}, \boldsymbol{\varphi}, \boldsymbol{\lambda} | y_t).$$

Properties of the maximum likelihood estimators (MLEs) have been studied in Andrews et al. (2006) and Lanne and Saikkonen (2011a). Both papers provide regularity conditions under which the MLE is consistent and asymptotically normal. The latter property requires in particular that $\mathbb{E}(|\varepsilon_t|^\delta) < \infty$ for some $\delta \geq 2$. For consistency, the condition on δ is less restrictive. In particular, consistency when $\delta \in (0, 1)$ has been shown in Gouriéroux and Zakoïan (2013).⁶

⁵The term ‘approximate’ stems from the fact that the sample used in the likelihood contains only $T - (r + s)$ terms. As shown in Breidt et al. (1991), this quantity is only an approximation of the true joint density of the data vector $\mathbf{y} = [y_1, \dots, y_T]'$.

⁶Proposition 2.1 in Gouriéroux and Zakoïan (2013) shows that assuming the summability condition $\sum_{j=-\infty}^{\infty} |\psi_j|^\delta < \infty$ for some $\delta \in (0, 1)$ ensures absolute summability of the MA coefficients, i.e., the convergence of the series y_t with probability 1. Then the process y_t is well defined for any *iid* sequence ε_t such that $\mathbb{E}(|\varepsilon_t|^\delta) < \infty$ for some $\delta \in (0, 1)$. This extension is extremely useful in the context where $\mathbb{E}(|\varepsilon_t|)$ is not defined, like in the Cauchy case.

In the econometric literature the Student's t -distribution (and the Cauchy as a particular case) has almost exclusively been used in applications, since it is a straightforward choice which allows to incorporate heavy tail behavior, often observed in financial data. Relatively little is known about the properties of the MLE of mixed causal-noncausal models if $f_\varepsilon(\varepsilon_t; \boldsymbol{\lambda})$ is not assumed to be known. Properties of the quasi-MLE (QMLE) are derived in Huang and Pawitan (2000) and Wu and Davis (2010) for the specific case of a Laplacian objective function, resulting in the least absolute deviation estimator (LAD). Under suitable assumptions (restrictions) on the underlying noise sequence ε_t , the LAD estimator is consistent and asymptotically normal. In the sequel we present and assess the Student's t MLE and the LAD estimator in more detail.

2.3.1 Student's t Maximum Likelihood Estimation

If ε_t is a sequence of *iid* zero mean random variables with probability density function

$$f_\varepsilon(\varepsilon_t | \sigma, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\sigma} \left(1 + \frac{1}{\nu} \left(\frac{\varepsilon_t}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}},$$

the corresponding (approximate) log-likelihood function with $\boldsymbol{\lambda} = [\sigma, \nu]'$, conditional on the observed data $\mathbf{y} = [y_1, \dots, y_T]'$ can be formulated as

$$l_y(\boldsymbol{\phi}, \boldsymbol{\varphi}, \alpha, \boldsymbol{\lambda} | \mathbf{y}) = (T-p) \left[\ln \left(\Gamma \left(\frac{\nu+1}{2} \right) \right) - \ln \left(\sqrt{\nu\pi\sigma^2} \right) - \ln \left(\Gamma \left(\frac{\nu}{2} \right) \right) \right] - \frac{\nu+1}{2} \sum_{t=r+1}^{T-s} \ln \left(1 + \frac{1}{\nu} \left(\frac{\phi(L)\varphi(L^{-1})y_t - \alpha}{\sigma} \right)^2 \right), \quad (2.10)$$

where $p = r+s$. The scale parameter is denoted by σ , $\nu > 0$ are degrees of freedom, and $\Gamma(\cdot)$ is the gamma function. Thus, the MLE corresponds to the solution of the problem:

$$\hat{\boldsymbol{\theta}}_{ML} = \arg \max_{\boldsymbol{\theta} \in \Theta} l_y(\boldsymbol{\theta} | \mathbf{y}),$$

with $\boldsymbol{\theta} = [\boldsymbol{\phi}', \boldsymbol{\varphi}', \alpha, \boldsymbol{\lambda}']'$ and Θ is a permissible parameter space containing the true value of $\boldsymbol{\theta}$, say $\boldsymbol{\theta}_0$, as an interior point. Since an analytical solution of the score function is not directly available, gradient based (numerical) procedures like

the Berndt-Hall-Hall-Hausman (BHHH) and Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithms can be used to find $\hat{\boldsymbol{\theta}}_{ML}$.

If $\nu > 2$, and hence $\mathbb{E}(|\varepsilon_t|^2) < \infty$, the MLE is \sqrt{T} -consistent and asymptotically normal. Denote the effective sample size $(T - p)$ used to compute the approximate MLE by n and define the $(n \times 1)$ series $\mathbf{u} \equiv \mathbf{U}_t = [u_{r+1}, \dots, u_{T-s}]'$ up to $\mathbf{U}_{t+s} = [u_{r+s+1}, \dots, u_T]'$, $\mathbf{V}_{t-r} = [v_1, \dots, v_{T-p}]'$ up to $\mathbf{v} \equiv \mathbf{V}_t = [v_{r+1}, \dots, v_{T-s}]'$ and $\boldsymbol{\varepsilon} = [\varepsilon_{r+1}, \dots, \varepsilon_{T-s}]'$. Then we construct the matrices $\mathbf{Z} = [\mathbf{U}_{t+1}, \dots, \mathbf{U}_{t+s}]$ and $\mathbf{Q} = [\mathbf{V}_{t-1}, \dots, \mathbf{V}_{t-r}]$, which are of dimensions $(n \times s)$ and $(n \times r)$ respectively. Using this notation, we can write the autoregressions defined in (2.6) and (2.7) in matrix notation as follows:

$$\mathbf{u} = \mathbf{Z}\boldsymbol{\varphi} + \boldsymbol{\varepsilon},$$

$$\mathbf{v} = \mathbf{Q}\boldsymbol{\phi} + \boldsymbol{\varepsilon}.$$

From this it follows (see e.g., Fonseca et al., 2008) that in case of the mixed causal-noncausal model

$$\sqrt{n}(\hat{\boldsymbol{\phi}}_{ML} - \boldsymbol{\phi}_0) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \frac{\nu+3}{\nu+1}\sigma^2\boldsymbol{\Upsilon}_{\boldsymbol{\varphi}}^{-1}\right), \quad (2.11)$$

$$\sqrt{n}(\hat{\boldsymbol{\varphi}}_{ML} - \boldsymbol{\varphi}_0) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \frac{\nu+3}{\nu+1}\sigma^2\boldsymbol{\Upsilon}_{\boldsymbol{\phi}}^{-1}\right), \quad (2.12)$$

holds. We use the notation $\boldsymbol{\Upsilon}_{\boldsymbol{\varphi}} = \mathbb{E}[\mathbf{Q}'\mathbf{Q}]$ and $\boldsymbol{\Upsilon}_{\boldsymbol{\phi}} = \mathbb{E}[\mathbf{Z}'\mathbf{Z}]$, where $\boldsymbol{\varphi}$ and $\boldsymbol{\phi}$ signify the relation between the unobserved values \mathbf{u} , \mathbf{v} and \mathbf{y} as defined in (2.6)-(2.7). These quantities can be estimated consistently by $(1/n)\sum_{i=1}^n \mathbf{Q}'_i\mathbf{Q}_i$ and $(1/n)\sum_{i=1}^n \mathbf{Z}'_i\mathbf{Z}_i$, where \mathbf{Q}_i [resp. \mathbf{Z}_i] denotes the i th row of the matrix \mathbf{Q} [resp. \mathbf{Z}]. For large ν , i.e., $\nu \rightarrow \infty$, l_y approaches the Gaussian (log)-likelihood, and the model parameters cannot be consistently estimated anymore. The derivation of this result is collected in Appendix 2.B; the equivalence of information sets underlying the approximate maximum likelihood estimation is considered in Appendix 2.C. In Section 2.4 we explore the properties of the MLE by means of simulations.

The limiting distribution of these parameters could also be considered in the infinite variance setting. Findings in Davis et al. (1992) and Andrews and Davis (2013) however show that a closed form solution for these distributions does not exist (and hence is intractable in practice). One could overcome this problem by means of bootstrapping and simulation-based methods. As our aim is to introduce a straightforward procedure to compute (for example) standard errors of paramete-

ters which is not based on computing the Hessian using gradient based (numerical) procedures, we do not consider the infinite variance case here.

2.3.2 Least Absolute Deviation Estimation

The LAD criterion is derived via a likelihood approximation assuming that the underlying noise ε_t is mean zero Laplacian, i.e., $f_\varepsilon(\varepsilon_t|\sigma) = \frac{1}{2\sigma} \exp(-|\varepsilon_t|/\sigma)$, with σ denoting the scale parameter. Then the approximate log-likelihood function is

$$l_y(\phi, \varphi, \alpha, \sigma | y_1, \dots, y_T) = -n \ln(2\sigma) - \frac{1}{\sigma} \sum_{t=r+1}^{T-s} |\phi(L)\varphi(L^{-1})y_t - \alpha|. \quad (2.13)$$

By maximizing (2.13) with respect to σ , we obtain the maximizer

$$\hat{\sigma} = n^{-1} \sum_{t=r+1}^{T-s} |\phi(L)\varphi(L^{-1})y_t - \alpha|,$$

which we can use to concentrate out σ in the same equation (2.13) to obtain the objective function

$$-n \left(1 + \ln(2n^{-1}) + \ln \left(\sum_{t=r+1}^{T-s} |\phi(L)\varphi(L^{-1})y_t - \alpha| \right) \right),$$

which is (given T) maximized if the criterion $\sum_{t=r+1}^{T-s} |\phi(L)\varphi(L^{-1})y_t - \alpha|$ is minimized. Minimizing this criterion with respect to ϕ, φ and α corresponds to a nonlinear program, which can be solved numerically with various approaches. We find the BFGS algorithm mentioned earlier to be a fast and stable variant.

Wu and Davis (2010) have shown⁷ that if ε_t is a sequence of *iid* random variables with mean zero, median zero, finite variance and probability density function $f_\varepsilon(\varepsilon_t; \boldsymbol{\lambda})$ that is continuous in a neighborhood of zero, the LAD estimators (minimizing the modified objective function above) are \sqrt{T} -consistent and asymptotically normal. Again, conditional on the formulation of the unobserved causal and noncausal components and by using standard results in the literature (e.g.,

⁷However, based on a slightly different formulation of the initial mixed causal-noncausal model (see also Lanne and Saikkonen (2011a, section 3) for the equivalence of both representations).

(Knight, 1998) we find that

$$\sqrt{n}(\hat{\phi}_{LAD} - \phi_0) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \frac{1}{4f_\varepsilon^2(0)} \mathbf{\Upsilon}_\varphi^{-1}\right), \quad (2.14)$$

$$\sqrt{n}(\hat{\varphi}_{LAD} - \varphi_0) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \frac{1}{4f_\varepsilon^2(0)} \mathbf{\Upsilon}_\phi^{-1}\right). \quad (2.15)$$

In order to use these two expressions for inference, we need to estimate the density of the errors at zero. We can use a logistic kernel $K(x) = \exp(-x) / [1 + \exp(-x)]^2$ and estimate $f_\varepsilon(0)$ as

$$\hat{f}_\varepsilon(0) = \frac{1}{nb_n} \sum_{t=r+1}^{T-s} K\left(\frac{\hat{\phi}_{LAD}(L)\hat{\varphi}_{LAD}(L^{-1})y_t - \hat{\alpha}_{LAD}}{b_n}\right),$$

where we choose the bandwidth $b_n = 0.9An^{-1/5}$, with $A = \min\left[std(\hat{\varepsilon}_t), \frac{iqr(\hat{\varepsilon}_t)}{1.34}\right]$, with *std* and *iqr* being the standard deviation and the interquantile range of the residuals, respectively (see Silverman, 1986).

2.4 Simulation Study

In this section, we assess how the asymptotic results derived in (2.11)-(2.12) and (2.14)-(2.15) behave in finite samples by means of Monte Carlo experiments. In particular, we are interested in the behavior of the estimators in ‘‘boundary’’ cases where critical assumptions are violated. This includes for example the case when the error distribution is ‘very close’ to a Gaussian one, or has infinite variance. The data are generated by a mixed causal-noncausal model with both lag and lead order equal to one:

$$(1 - \phi_1 L)(1 - \varphi_1 L^{-1})y_t = \varepsilon_t,$$

under varying assumptions on the error distribution and different sample sizes ($T = 200$ and $T = 800$). The number of replications is 10,000 in all simulation experiments. More specifically, we investigate the finite sample properties of the estimators under the assumption that the errors ε_t follow

1. a t -distribution with 10 degrees of freedom.
2. a t -distribution with 3 degrees of freedom. The skewness and kurtosis are consequently undefined.

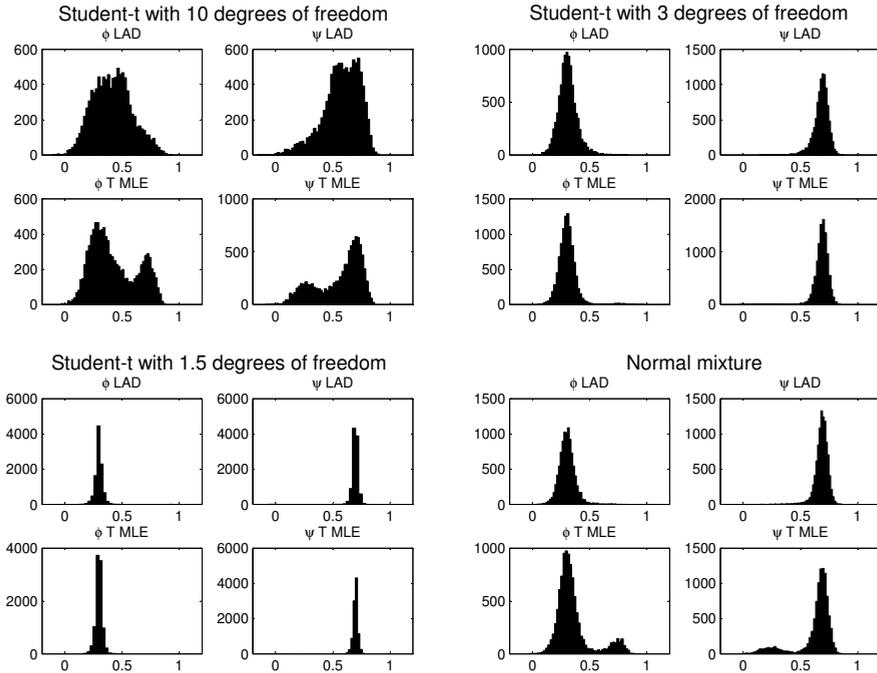


Figure 2.2: Histograms of estimators based on $T = 200$ observations

3. a t -distribution with 1.5 degrees of freedom. In this case the variance, the skewness and the kurtosis are undefined.
4. a normal mixture $\varepsilon_t = p_1x_1 + p_2x_2 + p_3x_3$ with equal probabilities $p_j = 1/3$, $j = 1, 2, 3$ and $x_1 \sim \mathcal{N}(-1, 2)$, $x_2 \sim \mathcal{N}(0, 0.5)$, $x_3 \sim \mathcal{N}(1, 2)$

To get a first impression of the behavior of $\hat{\phi}_1$ and $\hat{\varphi}_1$ we plot in Figure 2.2 for $T = 200$ and in Figure 2.3 for $T = 800$ the histograms of the estimators obtained in all simulation repetitions. The true parameter values are set to $\phi_1 = 0.3$ and $\varphi_1 = 0.7$. We can observe that the more the underlying distribution of the errors is approaching the normal distribution the less precise the resulting estimates. Moreover, the heavier the tails the faster the estimator seems to converge. The LAD seems to be slightly more accurate than the MLE, in particular when $T = 200$ for case 4 (asymmetric distribution). This implies that the identification of the mixed causal-noncausal model can be done for relatively small samples, for very fat-tailed distributions and without a reference to a well specified parametric distribution in a MLE approach such as a t -distribution.

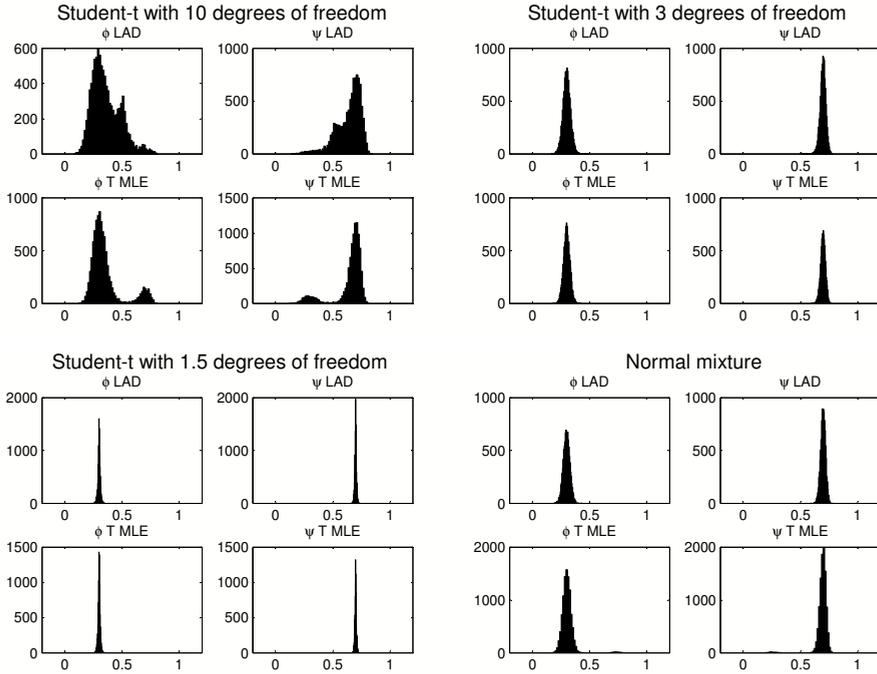
Figure 2.3: Histograms of estimators based on $T = 800$ observations

Table 2.1 summarizes the results of the simulation study in terms of means and standard deviations (SD) of the estimated parameters over all replications. The entry AD corresponds to the asymptotic standard deviations (based on specifications (2.11)-(2.12) and (2.14)-(2.15) computed as the average over the 10,000 replications. We see that for cases 2, 3 and 4 (i.e., all our specifications but the Student's $t(10)$ case) point estimates are very close to the true DGP. The only exception is the MLE for the mixture in finite samples. Both MLE and LAD present some bias for the t -distribution with 10 degrees of freedom. To the question: how fat must the tails of the error distribution be in practice, we conclude that the heavier the tails are, the more accurate the estimation of the parameters is. This is particularly true when using the LAD but also the Student's t MLE performs well. This is in line with results in Davis et al. (1992), who find that both the MLE and LAD estimator converge faster in the case of infinite variance autoregressive processes. When comparing the estimated standard error (SD) to the asymptotic one (AD) it emerges that the approximation is poor for $t(10)$ but very accurate for more leptokurtic and asymmetric distributions, in particular for

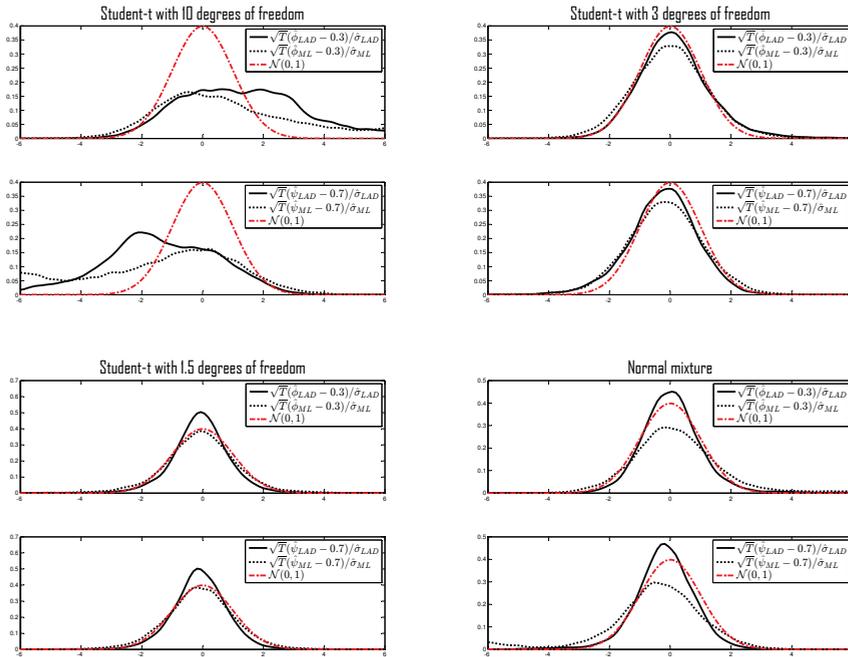


Figure 2.4: Density plots of the estimators's sampling distribution and their asymptotic limit based on $T = 200$ observations

the LAD estimator. This is illustrated in Figures 2.4 and 2.5 which show how fast the asymptotic approximations in (2.11)-(2.12) and (2.14)-(2.15) approach their distributional limit.

2.5 Empirical Applications

2.5.1 Empirical Illustration 1: Solar Panel Bubbles

We first investigate the existence of a bubble observed in the solar panels market in Belgium (more exactly the Walloon Region, i.e., the southern part of Belgium). Photovoltaic systems are technologies that use solar panels to directly convert the solar energy into electricity. This technology has been promoted in many countries, in Europe in particular, in order to reduce carbon dioxide emissions. While solar panels are nowadays relatively cheap due to the massive imports from China, this was not the case until roughly 2010, say. Thus, policy makers, for instance to

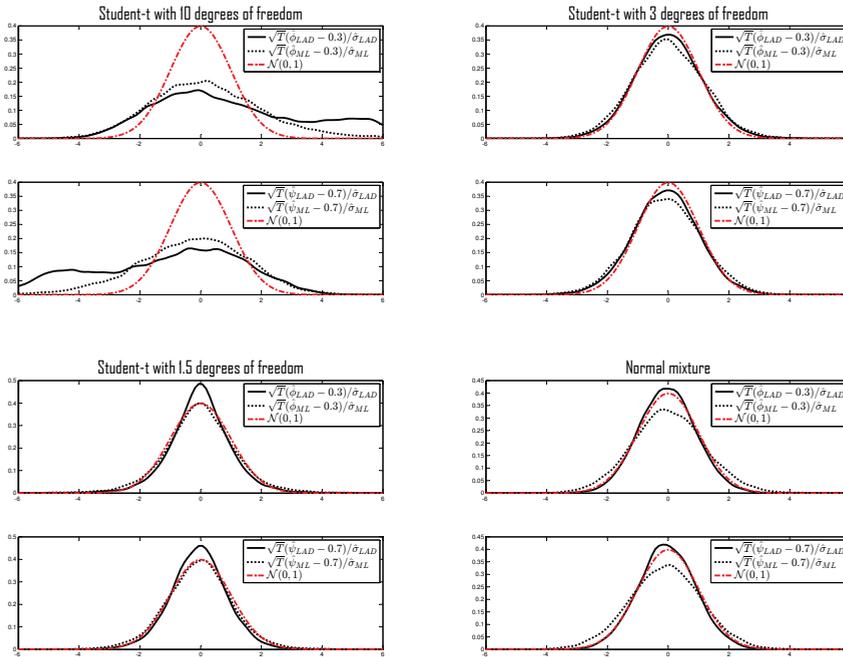


Figure 2.5: Density plots of the estimators's sampling distribution and their asymptotic limit based on $T = 800$ observations

meet the 1997 Kyoto Protocol that commits parties to reduce greenhouse gases emissions, have started to subsidize installation of such photovoltaic systems at the end of the 90's. This financial assistance has indeed helped to make this technology popular but has also created some side effects. Solar panels prices paid by final consumers did not tend to decrease as fast as expected in the short-run because some companies have maintained high prices, incorporating those financial helps in their bills to consumers.

Figure 2.6 displays the series we use on the number of solar panels installed (Nombre d'UPD, denoted $Units_t$ hereafter) and the power in kilovolt ampere (Puissances, denoted kVA_t hereafter) they generate. Monthly variables spanning 2008:01-2014:08 are released by the *Commission Wallone pour l'Energie* (CWaPE), the official organization that regulates the electricity and gas markets.⁸ The familiar bubble and burst pattern is obvious in both series. To some extent, the bubble could have been foreseen for several reasons. First, regional

⁸See <http://www.cwape.be/>

		$t(10)$		$t(3)$		$t(1.5)$		Mixture	
		$\hat{\phi}_{LAD}$	$\hat{\varphi}_{LAD}$	$\hat{\phi}_{LAD}$	$\hat{\varphi}_{LAD}$	$\hat{\phi}_{LAD}$	$\hat{\varphi}_{LAD}$	$\hat{\phi}_{LAD}$	$\hat{\varphi}_{LAD}$
Mean	$T = 200$	0.410	0.575	0.309	0.681	0.299	0.696	0.304	0.686
	$T = 800$	0.363	0.649	0.300	0.697	0.299	0.699	0.299	0.698
SD	$T = 200$	0.169	0.160	0.082	0.067	0.044	0.033	0.071	0.057
	$T = 800$	0.125	0.111	0.033	0.025	0.011	0.008	0.030	0.023
AD	$T = 200$	0.083	0.074	0.069	0.052	0.033	0.026	0.069	0.055
	$T = 800$	0.041	0.034	0.031	0.024	0.011	0.008	0.031	0.023
		$\hat{\phi}_{ML}$	$\hat{\varphi}_{ML}$	$\hat{\phi}_{ML}$	$\hat{\varphi}_{ML}$	$\hat{\phi}_{ML}$	$\hat{\varphi}_{ML}$	$\hat{\phi}_{ML}$	$\hat{\varphi}_{ML}$
Mean	$T = 200$	0.429	0.554	0.306	0.683	0.301	0.696	0.355	0.631
	$T = 800$	0.353	0.644	0.299	0.698	0.299	0.699	0.306	0.691
SD	$T = 200$	0.199	0.198	0.084	0.077	0.055	0.042	0.155	0.155
	$T = 800$	0.138	0.135	0.029	0.022	0.010	0.007	0.057	0.053
AD	$T = 200$	0.060	0.055	0.050	0.038	0.022	0.017	0.056	0.047
	$T = 800$	0.032	0.026	0.025	0.018	0.009	0.006	0.029	0.022

Table 2.1: Means and standard deviations of the MLE and LAD estimators

policy makers did not adapt the subsidies (greenhouse certificates) to the new market conditions (price decreased on the world market). Thus the internal return provided by the solar panels was much above what the financial market had to offer. The second problem was due to the financial crisis hitting Europe after 2008. This has imposed budget cuts such that the financial assistance had to stop. This development has been communicated far in advance. Consequently, this announcement has created a boom and then the burst phenomenon characteristic of such a bubble. It is described in the literature that regulatory changes give rise to a sharp decrease contrarily to expectations and sentiment reversals that produce a more gradual decrease (dot.com bubble in the US for instance).

Implementation of $MAR(r, s)$ Models

In order to identify mixed causal-noncausal models in those series we will follow the following strategy:

1. Lanne and Saikkonen (2011a) propose to estimate a conventional causal autoregressive process by OLS or Gaussian maximum likelihood and to determine the lag order p using information criteria (AIC, BIC or HQ for instance). Diagnostic tests can be performed to investigate whether the determined lag order adequately captures the autocorrelation in the error process. Simulation results (available upon request) favor the use of HQ.

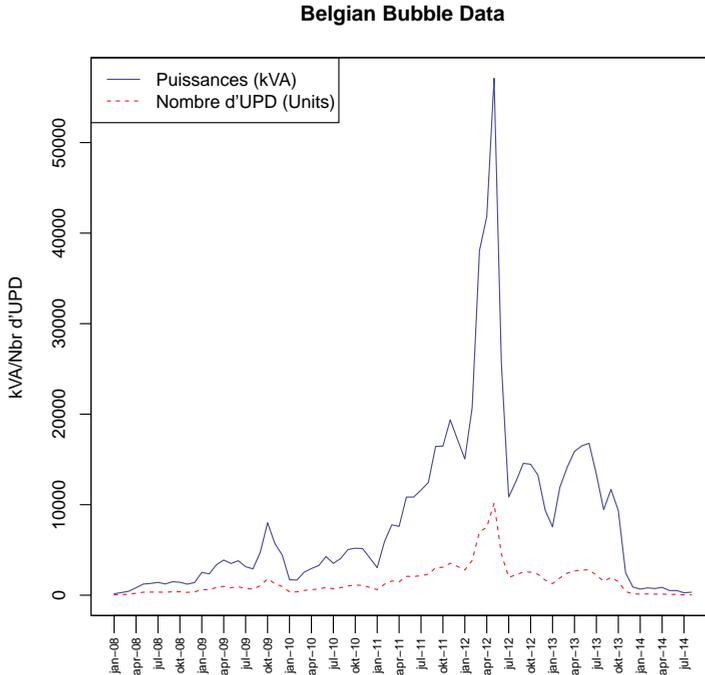


Figure 2.6: Number of solar panel units and power

2. We test for the null of normality on the residuals of the $AR(p)$ using e.g., the Jarque-Bera⁹ and one-sample Kolmogorov-Smirnov (KS) test. It should be noted, that in case the null hypothesis of normality is not rejected, there is no need to consider noncausal or mixed causal-noncausal models, as the causal and noncausal polynomial cannot be distinguished.
3. We test for *iid*-ness. If the true process (i.e., the DGP) is *iid* causal, then the errors from the $AR(p)$ must be *iid*. One should reject the *iid* null hypothesis if we estimate a causal model using data generated by an *iid* noncausal DGP. Several tests can be used (e.g., BDS test) but we propose for simplicity to look at the residuals from the causal $AR(p)$, denoted $\hat{\varepsilon}_t$, and to run a

⁹Strictly speaking, the Jarque-Bera test does not test for normality, but for the skewness and kurtosis of the data to match the normal distribution. We acknowledge this fact, but because of the close correspondence, we treat the Jarque-Bera test as a test for normality.

regression from $\hat{\varepsilon}_t$ on $\hat{\varepsilon}_{t-1}^2$ up to $\hat{\varepsilon}_{t-m}^2$:

$$\hat{\varepsilon}_t = \mu + \delta_1 \hat{\varepsilon}_{t-1}^2 + \dots + \delta_m \hat{\varepsilon}_{t-m}^2 + u_t, \quad (2.16)$$

and to test for *iid*-ness using the null hypothesis $H_0 : \delta_1 = \dots = \delta_m = 0$. This test is asymptotically $\chi_{(m)}^2$ distributed under normality. Because this is not the case here, its distribution can be tabulated. This is the natural extension to the procedure proposed by Gouriéroux and Zakoïan (2013) for $m = 1$ lag.

4. After we established the order p that captures the autocorrelation in the series and we find that non-normality is detected and the *iid*-ness of the causal residuals can be rejected, the next step is to select a model among all $MAR(r, s)$ specifications with $r + s = p$. To do so, we apply the Student's t MLE as well as the LAD estimators that we described in Section 3 for each potential model (within $p = r + s$); we take the one that gives the highest value for the likelihood. Note that every model has got the same number of regressors, hence the use of information criteria is superfluous here. Note that, although an intercept is included in our specifications (2.10) and (2.13), demeaning the series first (but leaving the intercept in the regression) seems to improve the results in terms of speed and accuracy of the convergence process.

MLE and LAD Results

Table 2.2 provides the $AR(p)$ models (estimated by OLS) for both series using usual BIC, the value of the Jarque-Bera (JB) normality test, the LM tests for the null hypothesis of no-autocorrelation and *iid*-ness. Those tests are obtained on residuals after having estimated $AR(1)$ processes. The notation $LM(m)$ for $m \geq 1$ indicates that m lagged residuals are used in the auxiliary regression equations testing for *iid*-ness (see (2.16) for explicit form) and white noise. The white noise null hypothesis is not rejected but both the normality and *iid* hypotheses are. The Augmented Dickey-Fuller test does not reject the null marginally (p -values of 0.06) but it is well known that the test suffers from both size and power distortions in this situation (see e.g., Gouriéroux and Zakoïan, 2013 or Saikkonen and Sandberg, 2016).

We observe that there is a strong rejection of the null of normality. This is a first indication that it is justified to investigate the presence of noncausality in

	Variables	
	$Units_t$	kVA_t
<i>BIC</i>	AR(1)	AR(1)
<i>JB</i> : $H_0 = normality$	727.76	748.21
<i>LM</i> (1) : $H_0 = white\ noise$	0.00	0.03
<i>LM</i> (4) : $H_0 = white\ noise$	1.17	1.11
<i>LM</i> (1) : $H_0 = iid$	19.18	19.62
<i>LM</i> (4) : $H_0 = iid$	9.78	9.38
<i>ADF - test</i>	-2.78	-2.77
Log-likelihood - LAD		
MAR(1,0)	-680.071	-817.068
MAR(0,1)	-678.725*	-814.834*
Log-likelihood - MLE		
MAR(1,0)	-593.189	-724.129
MAR(0,1)	-588.327*	-718.403*

Table 2.2: Summary Statistics

the series of interest. Using BIC, an AR(1) is selected for both series. We will thus consider either a purely causal MAR(1,0) or a purely noncausal MAR(0,1). Table 2.2 also reports the value of the log-likelihood and we see that for both series a noncausal MAR(0,1) is preferred for $p = 1$, as this model maximizes the log-likelihood for both the MLE and LAD estimation method.

The two MAR(0,1) equations are estimated both by maximum likelihood and LAD. The 95% confidence intervals are in brackets below estimated parameters. There are no large differences whether one takes the number of solar panels or the power they generate, but we observe that the parameters estimated by LAD are more persistent such that the confidence intervals marginally include the unit root case. We find the MLE results more plausible although the small number of observations does not allow to draw sharp conclusions; the estimated degrees of freedom tell us that we are close to the Cauchy distribution for the error process.

$$\text{Student's } t \text{ MLE} \left\{ \begin{array}{l} \widehat{Units}_t = \begin{array}{l} 0.928 \\ [0.893, 0.964] \end{array} Units_{t+1}, \quad \hat{\nu} = 1.09 \\ \widehat{kVA}_t = \begin{array}{l} 0.907 \\ [0.878, 0.967] \end{array} kVA_{t+1}, \quad \hat{\nu} = 0.98 \end{array} \right.$$

$$\text{LAD} \left\{ \begin{array}{l} \widehat{Units}_t = \underset{[0.908, 1.007]}{0.957} Units_{t+1}, \\ \widehat{kVA}_t = \underset{[0.928, 1.010]}{0.968} kVA_{t+1}, \end{array} \right.$$

An additional outcome that is out of the scope of this chapter is whether the econometrician could have anticipated the crash of the bubble. Gouriéroux and Jasiak (2016) propose such a risk measure and proceed in the following way. One cuts off the data from the crash onwards, estimates an $MAR(r, s)$ on the remaining observations, say (y_1, \dots, y_{T^*}) , and uses these results to derive a joint predictive density for future values $(y_{T^*+1}, \dots, y_{T^*+H})$. From these predictive densities probabilities can be computed for different types of scenarios like e.g., $y_{T^*} < y_{T^*+1} < y_{T^*+2}$, which signifies the probability that the bubble keeps on increasing in the upcoming two time periods (given the last data point y_{T^*}). Our sample is too small to realize this exercise.

2.5.2 Empirical Illustration 2: Realized Volatilities

The previous application was based on a very small sample size, around 75 data points. This can affect the outcomes in two ways. First, we have seen in Table 2.1 that results are obviously more accurate for larger samples. A second opposite effect is that the bubble phenomena that we can clearly observe in small samples (both in simulated graphs of Section 2 and on Belgian solar panels data) can be considered as a zoom of a larger process. When one constructs a series with thousands of observations, we stack to some extent those processes next to each other and the concatenated series generated/observed on a larger sample looks more random. In this second application we identify mixed causal-noncausal models in daily realized variances for almost 4,000 observations. To the best of our knowledge, this is the first time that these type of models are used on this kind of data. There are no doubts that Gaussianity as well as *iid*-ness is rejected for all series. Hence, we do not repeat steps 2 and 3 of the estimation strategy outlined in Section 2.5.1.

The realized variances are computed using

$$RV_t \equiv \sum_{i=1}^M r_{t,i}^2,$$

where $r_{t,i}$ are the high frequency intra-day returns, observed for M intra-day periods each day. For instance $M = 288$ for 5-min returns on 24 hours or around 80 when the market is open between 9am to 4pm. In order to attenuate the impact of jumps, we work in this chapter with bipower variation (BV henceforth) series instead of realized variances (Barndorff-Nielsen and Shephard, 2004). They are computed as (Bauwens et al., 2012)

$$BV_t \equiv \frac{\pi}{2} \left(\frac{M}{M-1} \right) \sum_{i=2}^M |r_{t,i}| |r_{t,i-1}|.$$

The bipower variations also provide a consistent measure of the integrated volatility associated with standard theoretical continuous time diffusion models. However, BVs are designed to be more robust to jumps than RVs because they are computed on products between two consecutive returns instead of the squared returns. We also consider in the application the median realized volatility measure (see Andersen et al., 2010)

$$Med-RV_t \equiv \frac{\pi}{6 - 4\sqrt{3} + \pi} \left(\frac{M}{M-2} \right) \sum_{i=2}^{M-1} \text{med} (|r_{t,i}| |r_{t,i-1}| |r_{t,i+1}|)^2.$$

Our series are obtained from 5 minute data and are from Oxford-Man Institute of Quantitative Finance, Library version 0.2 (series code bv5 & medrv).¹⁰ We use the period 03/01/2000 to 08/10/2014 (i.e., $T = 3858$ observations¹¹) for twenty-one equity indexes whose names are given in Table 2.3.

In this analysis we have replaced for simplicity the missing days, that are due to closing market, by the average of the series over the whole period. Alternative approaches can obviously be used. We have not divided the sample into different subperiods and hence the financial crisis is also included in our samples. As an example, Figure 2.7 shows the computed Med-RV and BV series for the equity indexes of S&P500, FTSE, Nikkei and DAX.

The heterogenous autoregressive model (HAR hereafter) proposed by Corsi (2009) is commonly used to approximate the long memory features of realized volatilities on similar series (see, Corsi et al., 2012). The HAR for the RV_t variable

¹⁰See Heber et al. (2009).

¹¹We do not consider the first 600 observations that are not available for the Canadian index.

Ox.Man name	Common name	Country	MAR(r, s)-BV	MAR(r, s)-Med RV
AEX	AEX Index	Netherlands	MAR(6,2)	MAR(7,1)
AORD2	All Ordinaries	Australia	MAR(8,2)	MAR(2,7)
BVSP	Bovespa Index	Brazil	MAR(10,4)	MAR(12,6)
DJI2	DJIA	USA	MAR(7,5)	MAR(8,3)
FCHI2	CAC 40	France	MAR(9,0)	MAR(5,3)
FTSE2	FTSE 100	UK	MAR(7,2)	MAR(5,3)
FTSEMIB	FTSE MIB	Italy	MAR(5,1)	MAR(6,3)
FTSTI	FT straits time index	Singapore	MAR(6,15)	MAR(0,16)
GDAXI2	DAX	Germany	MAR(12,2)	MAR(12,2)
GSPTSE	S&P/TSX Composite Index	Canada	MAR(14,13)	MAR(15,13)
HSENG	Hang Seng	Asia	MAR(12,1)	MAR(17,1)
IBEX2	IBEX 35	Spain	MAR(5,1)	MAR(5,1)
IXIC2	Nasdaq 100	USA	MAR(4,7)	MAR(8,3)
KS11	KOSPI Composite Index	South Korea	MAR(5,2)	MAR(6,1)
MXX	IPC Mexico	Mexico	MAR(11,0)	MAR(6,2)
N2252	Nikkei 225	Japan	MAR(17,1)	MAR(12,6)
NSEI	S&P CNX Nifty	India	MAR(5,1)	MAR(1,1)
RUT2	Russell 2000	USA	MAR(9,5)	MAR(4,7)
SPX2	S&P 500	USA	MAR(6,7)	MAR(10,1)
SSMI	Swiss Market Index	Switzerland	MAR(7,2)	MAR(7,6)
STOXX50E	Euro STOXX 50	Eurozone	MAR(12,4)	MAR(10,2)

Table 2.3: Equity Indexes and MAR(r, s) identification

is such that

$$RV_t^{(d)} = \beta_0 + \beta^{(d)} RV_{t-1d}^{(d)} + \beta^{(w)} RV_{t-1d}^{(w)} + \beta^{(m)} RV_{t-1d}^{(m)} + v_t, \quad t = 1, 2, \dots, T,$$

where (d) , (w) , and (m) denote, respectively, the influence from yesterday and moving average impacts of last week average (assuming 5 days a week) and last month average (assuming 20 days a month) with $RV_t^{(w)} = \frac{1}{5} \sum_{j=0}^4 RV_{t-jd}^{(d)}$ and $RV_t^{(m)} = \frac{1}{20} \sum_{j=0}^{19} RV_{t-jd}^{(d)}$. Hence, this HAR is also a restricted form of a causal AR(20).

Developing a similar framework for mixed causal-noncausal models is out of the scope of this chapter, in particular the multiplicative nature of the MAR(r, s) necessitates to adapt the HAR structure. However we want to allow for a dependence on the past and the future at rather long displacements, namely at least one month in both time directions. We start by estimating a causal AR model by OLS for each series allowing a p_{\max} of 50 days. We use BIC to determine the lag order, even though it is known that information criteria probably overestimate the number of lags for highly volatile series (see e.g., Hecq, 1996). Depending on the indexes, p ranges from 2 to 28. Next, we search using the Student's t MLE approach, within $p = r + s$, the MAR(r, s) model that maximizes the log-

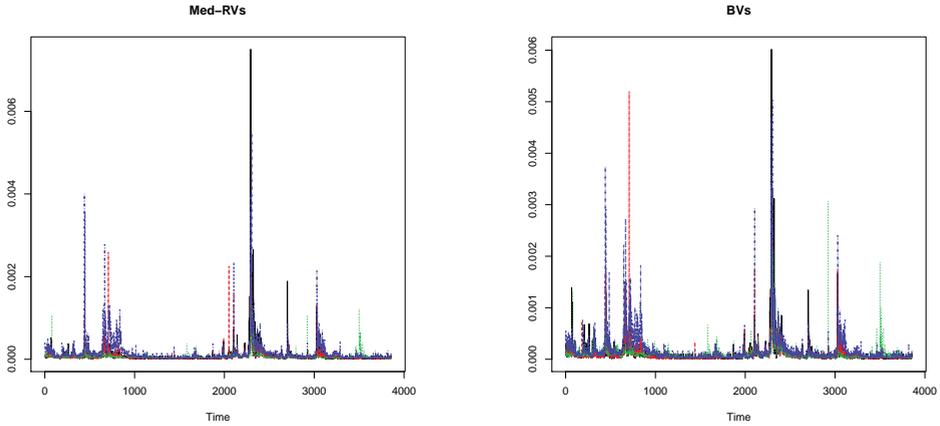


Figure 2.7: Med-RVs and BVs for S&P500, FTSE, Nikkei and DAX

likelihood. The last two columns of Table 2.3 report those $\text{MAR}(r, s)$ models for both volatility indicators. There are only two entries (resp. one entry) out of 42 series for which we get a pure causal (resp. purely noncausal) model. In most situations mixed causal-noncausal models are selected. We find degrees of freedom for the t -distribution between 1.1 and 2 for almost every series in our dataset. We leave the value added of considering the additional noncausal part for forecasting volatilities for further research, but this is clearly an important issue. Also the interaction between $\text{MAR}(r, s)$ models and time-varying volatilities is not considered here.

2.6 Conclusion

We do not claim that noncausality is everywhere. However we must recognize that the Gaussianity assumption is rejected in many macroeconomic and financial time series. This gives the opportunity to extend the usual Box-Jenkins approach in three different ways: (i) y_t is assumed to have a two-sided moving average representation, including both a causal component (current and lagged values of the disturbance term) and a noncausal component (future values of the error term) (ii) The error term is strong white noise, instead of weak white noise; (iii) The distribution of the error term can be leptokurtic, with the possibility of having infinite variance or even infinite expectation.

This chapter compares the finite sample performances of the MLE, assuming a t -distribution, with the LAD estimator. We show in a simulation study that both methods capture the causal and/or noncausal dependency particularly well when the distribution has very fat tails. These results are in line with findings in Davis et al. (1992) who find that both the MLE and LAD estimator converge faster for causal autoregressive processes with infinite variance than their counterparts with finite variance. The simulation study in this chapter shows that these results hold for mixed causal-noncausal processes as well.

In the empirical section we investigate two types of data: monthly solar panel data and daily realized volatility measures. In both applications, the presence of a noncausal component is detected (in most cases) and the estimated degrees of freedom for the t -distribution are found to be very small. Hence, the data confirms the interest in mixed causal-noncausal processes with potentially infinite variance or even expectation. The results of the MLE and LAD estimator are found to be very similar.

2.A Weak White Noise

It is a well-known phenomenon that if y_t is a noncausal AR(p) driven by an *iid* sequence of disturbances with mean zero and variance σ^2 , then y_t can also be expressed as a causal AR(p) driven by a newly created white noise sequence with mean zero and variance $\tilde{\sigma}^2$ (see e.g., Brockwell and Davis, 1991). In this appendix, we illustrate that this new sequence of innovations is generally not a strong white noise.

Similar to Brockwell and Davis (1991), we introduce a noncausal specification as an autoregressive model, for instance an AR(1), with roots inside the unit circle

$$y_t = \phi_1 y_{t-1} + \varepsilon_t,$$

where ε_t is a non-Gaussian *iid*($0, \sigma^2$) noise and $|\phi_1| > 1$. Note that, even though this model seems causal in terms of y_t , it has a noncausal stationary solution in terms of ε_t . That is, by rearranging terms and repeated substitution, it indeed follows that y_t admits the following moving average representation¹²

$$y_t = - \sum_{j=1}^{\infty} \phi_1^{-j} \varepsilon_{t+j}$$

with

$$\begin{aligned} \gamma_y(h) &= \text{Cov}(y_{t+h}, y_t) \\ &= \text{Cov} \left(- \sum_{j=1}^{\infty} \phi_1^{-j} \varepsilon_{t+h+j}, - \sum_{j=1}^{\infty} \phi_1^{-j} \varepsilon_{t+j} \right) = \sigma^2 \sum_{j=1}^{\infty} \phi_1^{-j} \phi_1^{-(h+j)} \\ &= \sigma^2 \phi_1^{-h} \sum_{j=1}^{\infty} \phi_1^{-2j} = \sigma^2 \phi_1^{-h} \frac{\phi_1^{-2}}{(1 - \phi_1^{-2})} = \sigma^2 \frac{1}{\phi_1^{|h|} (\phi_1^2 - 1)}. \end{aligned}$$

Let us define a new sequence $\varepsilon_t^* = y_t - \frac{1}{\phi_1} y_{t-1}$. It is straightforward (solving exercise 3.8 in Brockwell and Davis, 2002) that

$$\gamma_{\varepsilon^*}(h) = \text{Cov}(\varepsilon_{t+h}^*, \varepsilon_t^*)$$

¹²Complete iteration would yield $y_t = \phi_1^{-T} y_{t+T} - \sum_{j=1}^{\infty} \phi_1^{-j} \varepsilon_{t+j}$. However, since $|\phi_1| > 1$, it follows that $|\phi_1^{-1}| < 1$ and thus y_t is stationary. Then $\|y_t\|^2 = \mathbb{E}(y_t^2)$ is constant so that $\|y_t - (-\sum_{j=1}^{\infty} \phi_1^{-j} \varepsilon_{t+j})\|^2 = \phi_1^{-2T} \|y_{t+T}\|^2 \rightarrow 0$ as $T \rightarrow \infty$. Hence, it is reasonable to omit the value $\phi_1^{-T} y_{t+T}$ in our future computations.

$$= \gamma_y(h) - \phi_1^{-1}\gamma_y(h+1) - \phi_1^{-1}\gamma_y(h-1) + \phi_1^{-2}\gamma_y(h).$$

That is, $\gamma_{\varepsilon^*}(h) = 0$ for $h \neq 0$ and $\gamma_{\varepsilon^*}(h) = \frac{\sigma^2}{\phi_1^2}$ for $h = 0$, which confirms that ε_t^* is a white noise process. However, ε_t^* is not a *strong* white noise in general. In particular, we find that

$$\begin{aligned} \text{Cov}(\varepsilon_t^*, y_{t-1}^2) &= \text{Cov}\left(y_t - \frac{1}{\phi_1}y_{t-1}, y_{t-1}^2\right) \\ &= \mathbb{E}\left(-\sum_{j=1}^{\infty} \phi_1^{-j} \phi_1^{-2(j+1)} \varepsilon_{t+j}^3 + \frac{1}{\phi_1} \sum_{j=1}^{\infty} \phi_1^{-j} \phi_1^{-2j} \varepsilon_{t+j}^3\right) \\ &= \mu_3 \left(-\frac{1}{\phi_1^2} \sum_{j=1}^{\infty} \phi_1^{-j} \phi_1^{-2j} + \frac{1}{\phi_1} \sum_{j=1}^{\infty} \phi_1^{-j} \phi_1^{-2j}\right) \\ &= \mu_3 \left(-\frac{1}{\phi_1^2} \frac{\phi_1^{-3}}{(1-\phi_1^{-3})} + \frac{1}{\phi_1} \frac{\phi_1^{-3}}{(1-\phi_1^{-3})}\right) \\ &= \frac{\mu_3}{\phi_1^2(\phi_1^2 + \phi_1 + 1)}, \end{aligned}$$

which is unequal to zero if $\mathbb{E}(\varepsilon_t^3) = \mu_3 \neq 0$. Hence, if the original sequence of disturbances ε_t has a third moment different from zero, ε_t^* is no longer *iid*. If $\mu_3 = 0$, this does not imply independence. In fact, one should check $\text{Cov}(\varepsilon_t^*, y_{t-1}^j)$ for $j > 2$ as well to investigate dependence on the past. As we do not pursue to give a complete characterization but rather an illustration, we do not consider this exercise here. In particular, it is found that a sufficient and necessary condition to ensure *iid*-ness of ε_t^* is that ε_0 is Gaussian (see Breidt and Davis, 1992, for a formal proof).

2.B Proof Covariance Matrix

Let $\mathbf{y} = [y_1, \dots, y_n]'$ be a $(n \times 1)$ vector of observations and let $x_{j,i}$ be the j th exogenous variable out of a set of q variables observed at time i . Then we can construct a $(p \times 1)$ vector $\mathbf{x}_i = [x_{1,i}, \dots, x_{p,i}]'$ of all p exogenous variables at time i . Consequently, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]'$ is the $(n \times p)$ matrix of explanatory variables and $\boldsymbol{\varepsilon} = [\varepsilon_1, \dots, \varepsilon_n]'$ a $(n \times 1)$ vector of disturbances. We consider a model of the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.17}$$

where $\varepsilon_1, \dots, \varepsilon_n \stackrel{iid}{\sim} t(\sigma, \nu)$. We define $\boldsymbol{\theta}$ as the combination of the vector of regression parameters and distributional parameters, i.e. $\boldsymbol{\theta} = [\boldsymbol{\beta}', \sigma, \nu]' \in \mathbb{R}^p \times (0, \infty)^2$. Then the likelihood can be formulated as follows:

$$L(\boldsymbol{\beta}, \sigma, \nu | \mathbf{y}, \mathbf{X}) = \frac{\Gamma(\frac{\nu+1}{2})^n \nu^{n\nu/2}}{\Gamma(\frac{\nu}{2})^n \pi^{n/2} \sigma^n} \prod_{i=1}^n \left(\nu + \left(\frac{y_i - \mathbf{x}'_i \boldsymbol{\beta}}{\sigma} \right)^2 \right)^{-\frac{\nu+1}{2}}. \quad (2.18)$$

In order to prove that

$$-\mathbb{E} \left(\frac{\partial^2}{\partial \boldsymbol{\beta}' \partial \boldsymbol{\beta}} \ln L(\boldsymbol{\beta}, \sigma, \nu | \mathbf{y}, \mathbf{X}) \mid \boldsymbol{\beta} \right) = \sigma^{-2} \frac{\nu+1}{\nu+3} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}'_i,$$

we invoke Lemma A1 of Fonseca et al. (2008).

Lemma (Fonseca, Ferreira and Migon, 2008)

Consider the regression model $y \sim t(\nu, x' \boldsymbol{\beta}, \sigma^2)$. Define

$$c(\nu, \sigma^2) = \frac{\Gamma(\frac{\nu+1}{2}) \nu^{\nu/2}}{\Gamma(\frac{\nu}{2}) \sqrt{\pi \sigma^2}},$$

$z = \frac{y - x' \boldsymbol{\beta}}{\sigma}$, $a_k = \mathbb{E}[(\nu + z^2)^{-k}]$, $b_q = \mathbb{E}[(y - x' \boldsymbol{\beta})^q (\nu + z^2)^{-1}]$ and $c_q = \mathbb{E}[(y - x' \boldsymbol{\beta})^q (\nu + z^2)^{-2}]$. Then it holds that

$$(i) \quad \mathbb{E}(z) = 0 \text{ for } \nu > 1, \quad \mathbb{E}(z^2) = \frac{\nu}{\nu-2} \text{ for } \nu > 2, \quad \mathbb{E}(z^3) = 0 \text{ for } \nu > 3, \quad \mathbb{E}(z^4) = \frac{3\nu^2}{(\nu-2)(\nu-4)}.$$

$$(ii) \quad a_k = \mathbb{E}[(\nu + z^2)^{-k}] = \int_{-\infty}^{\infty} c(\nu, \sigma^2) (\nu + z^2)^{-\frac{\nu+2k+1}{2}} dy = \left(\frac{\nu+2k}{\nu} \right)^{\frac{\nu+2k+1}{2}} \frac{c(\nu, \sigma^2)}{c(\nu+2k, \frac{\nu\sigma^2}{\nu+2k})}.$$

$$\text{In particular, } a_1 = \frac{1}{\nu+1} \text{ and } a_2 = \frac{\nu+2}{\nu(\nu+1)(\nu+3)}.$$

$$(iii) \quad b_q = \mathbb{E}[(y - x' \boldsymbol{\beta})^q (\nu + z^2)^{-1}] = \int_{-\infty}^{\infty} (y - x' \boldsymbol{\beta})^q c(\nu, \sigma^2) (\nu + z^2)^{-\frac{\nu+3}{2}} = a_1 \mathbb{E}[(y - x' \boldsymbol{\beta})^q] \text{ where } y \sim t(\nu+2, x' \boldsymbol{\beta}, \frac{\nu\sigma^2}{\nu+2}).$$

$$(iv) \quad c_q = \mathbb{E}[(y - x' \boldsymbol{\beta})^q (\nu + z^2)^{-2}] = \int_{-\infty}^{\infty} (y - x' \boldsymbol{\beta})^q c(\nu, \sigma^2) (\nu + z^2)^{-\frac{\nu+5}{2}} = a_2 \mathbb{E}[(y - x' \boldsymbol{\beta})^q] \text{ where } y \sim t(\nu+4, x' \boldsymbol{\beta}, \frac{\nu\sigma^2}{\nu+4}).$$

The auxiliary facts in the lemma make it possible to rewrite the likelihood in (2.18) as follows:

$$L(\boldsymbol{\beta}, \sigma, \nu | \mathbf{y}, \mathbf{X}) = [c(\nu, \sigma^2)]^n \prod_{i=1}^n (\nu + z_i^2)^{-\frac{\nu+1}{2}}. \quad (2.19)$$

In a next step, we can calculate the corresponding log-likelihood, denoted as $\ell(\cdot)$, by applying the natural logarithmic transformation (which is monotonic):

$$\begin{aligned}\ell(\boldsymbol{\beta}, \sigma, \nu | \mathbf{y}, \mathbf{X}) &= n \ln[c(\nu, \sigma^2)] + \ln \left(\prod_{i=1}^n (\nu + z^2)^{-\frac{\nu+1}{2}} \right), \\ &= n \ln[c(\nu, \sigma^2)] - \frac{\nu+1}{2} \sum_{i=1}^n \ln(\nu + z^2), \\ &= n \ln[c(\nu, \sigma^2)] - \frac{\nu+1}{2} \sum_{i=1}^n \ln \left(\nu + \frac{y_i^2 - 2y_i \mathbf{x}'_i \boldsymbol{\beta} + \boldsymbol{\beta}' \mathbf{x}_i \mathbf{x}'_i \boldsymbol{\beta}}{\sigma^2} \right),\end{aligned}$$

where we wrote out z in the last equation, as we are interested in the derivative of the log-likelihood function with respect to $\boldsymbol{\beta}$. This derivative is given as

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\beta}, \sigma, \nu | \mathbf{y}, \mathbf{X})}{\partial \boldsymbol{\beta}} &= -\frac{\nu+1}{2} \sum_{i=1}^n (\nu + z^2)^{-1} \left(\frac{-2y_i \mathbf{x}'_i + 2\mathbf{x}_i \mathbf{x}'_i \boldsymbol{\beta}}{\sigma^2} \right), \\ &= \sum_{i=1}^n (\nu + z^2)^{-1} (\nu + 1) \left(\frac{\mathbf{x}_i (y_i - \mathbf{x}'_i \boldsymbol{\beta})}{\sigma^2} \right).\end{aligned}$$

Accordingly, the second derivative with respect to $\boldsymbol{\beta}$ can be computed. Applying the product and chain rule, we obtain:

$$\begin{aligned}\frac{\partial^2 \ell(\boldsymbol{\beta}, \sigma, \nu | \mathbf{y}, \mathbf{X})}{\partial \boldsymbol{\beta}' \partial \boldsymbol{\beta}} &= 2(\nu+1)\sigma^{-4}(\nu+z^2)^{-2} \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}'_i \right) \left(\sum_{i=1}^n y_i - \sum_{i=1}^n \mathbf{x}'_i \boldsymbol{\beta} \right)^2 \\ &\quad - (\nu+1)\sigma^{-2}(\nu+z^2)^{-1} \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}'_i \right).\end{aligned}$$

In order to obtain the Fisher information $\mathbb{I}(\cdot)$ with respect to $\boldsymbol{\beta}$, we have to take the expectation of this expression and multiply it by -1. That is,

$$\begin{aligned}\mathbb{I}(\boldsymbol{\beta}) &= -\mathbb{E} \left(\frac{\partial^2 \ell(\boldsymbol{\beta}, \sigma, \nu | \mathbf{y}, \mathbf{X})}{\partial \boldsymbol{\beta}' \partial \boldsymbol{\beta}} \middle| \boldsymbol{\beta} \right), \\ &= \mathbb{E} \left(-2(\nu+1)\sigma^{-4}(\nu+z^2)^{-2} \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}'_i \right) \left(\sum_{i=1}^n y_i - \sum_{i=1}^n \mathbf{x}'_i \boldsymbol{\beta} \right)^2 \right) \\ &\quad + \mathbb{E} \left((\nu+1)\sigma^{-2}(\nu+z^2)^{-1} \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}'_i \right) \right),\end{aligned}$$

$$\begin{aligned}
 &= -2(\nu + 1)\sigma^{-4} \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}'_i \right) \mathbb{E} \left[\left(\sum_{i=1}^n y_i - \sum_{i=1}^n \mathbf{x}'_i \boldsymbol{\beta} \right)^2 (\nu + z^2)^{-2} \right] \\
 &\quad + (\nu + 1)\sigma^{-2} \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}'_i \right) \mathbb{E}[(\nu + z^2)^{-1}].
 \end{aligned}$$

Using part (ii) and (iv) of the lemma, we can substitute expressions in for the two expectations in the last expression. We then arrive at

$$\begin{aligned}
 \mathbb{I}(\boldsymbol{\beta}) &= -2(\nu + 1)\sigma^{-4} \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}'_i \right) a_2 \mathbb{E} \left[\left(\sum_{i=1}^n y_i - \sum_{i=1}^n \mathbf{x}'_i \boldsymbol{\beta} \right)^2 \right] \\
 &\quad + (\nu + 1)\sigma^{-2} \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}'_i \right) a_1.
 \end{aligned}$$

Note that, since we invoked part (iv) of the lemma, $y \sim t(\nu + 4, x' \boldsymbol{\beta}, \frac{\nu \sigma^2}{\nu + 4})$ for the process y in the expectation. Hence, to obtain $\mathbb{E}(z^2)$, we have to divide the quantity by its new scale parameter $\frac{\nu \sigma}{\nu + 4}$. This leads to:

$$\begin{aligned}
 \mathbb{I}(\boldsymbol{\beta}) &= -2(\nu + 1)\sigma^{-2} \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}'_i \right) \frac{(\nu + 2)(\nu + 4)}{(\nu + 1)(\nu + 3)} \mathbb{E} \left[\frac{(\sum_{i=1}^n y_i - \sum_{i=1}^n \mathbf{x}'_i \boldsymbol{\beta})^2}{\nu \sigma^2 / (\nu + 4)} \right] \\
 &\quad + (\nu + 1)\sigma^{-2} \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}'_i \right) \frac{1}{\nu + 1}.
 \end{aligned}$$

The remaining expectation can be evaluated using part (i) of the lemma where $\mathbb{E}(z^2) = \frac{\nu + 4}{\nu + 2}$, since $y \sim t(\nu + 4, x' \boldsymbol{\beta}, \frac{\nu \sigma^2}{\nu + 4})$. That is, take $(\nu + 4)$ instead of ν in the expression for $\mathbb{E}(z^2)$. This gives us

$$\mathbb{I}(\boldsymbol{\beta}) = -2(\nu + 1)\sigma^{-2} \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}'_i \right) \frac{\nu + 2}{(\nu + 3)(\nu + 4)} \frac{\nu + 4}{\nu + 2} + \sigma^{-2} \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}'_i \right).$$

By collecting terms and simplifying expressions, we obtain:

$$\begin{aligned}
 \mathbb{I}(\boldsymbol{\beta}) &= \sigma^{-2} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}'_i \left(1 - 2 \frac{1}{\nu + 3} \right), \\
 &= \sigma^{-2} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}'_i \left(\frac{\nu + 1}{\nu + 3} \right).
 \end{aligned}$$

In order to make use of the results in Fonseca et al. (2008), we have to slightly adapt the model setup. That is, we consider

$$\phi(L)\varphi(L^{-1})y_t = \varepsilon_t, \quad (2.20)$$

where $\varepsilon_t \stackrel{iid}{\sim} t(\nu, \sigma)$. To ensure a similar model setup, we make use of the unobserved causal and noncausal components of the process y_t as defined in (2.6)-(2.7).

Let r and s denote the degrees of the polynomials $\phi(L)$ and $\varphi(L^{-1})$ respectively and let $n \equiv (T - p)$. We write

$$\mathbf{u} = \mathbf{Z}\varphi + \boldsymbol{\varepsilon}, \quad (2.21)$$

$$\mathbf{v} = \mathbf{Q}\phi + \boldsymbol{\varepsilon}, \quad (2.22)$$

where \mathbf{u} and \mathbf{v} are $(n \times 1)$ vectors of observations and \mathbf{Z} , \mathbf{Q} are $(n \times s)$ and $(n \times r)$ matrices as defined in Section 2.3.1. By reformulating the mixed causal-noncausal model in this way, the autoregressions (2.21) and (2.22) take a similar form as the original model (2.17), which makes it possible to apply the theory of Fonseca et al. (2008) immediately to these models.

2.C Equivalence of Information Sets

We consider the $MAR(r, s)$ model with unobserved causal and noncausal components u_t and v_t defined as in (2.6) and (2.7) in order to show the equivalence of various information sets in the following proposition.

Proposition

For an $MAR(r, s)$ model, the following information sets are equivalent:

- (i) (y_1, \dots, y_T)
- (ii) $(y_1, \dots, y_r, u_{r+1}, \dots, u_T)$
- (iii) $(v_1, \dots, v_{T-s}, y_{T-s+1}, \dots, y_T)$
- (iv) $(y_1, \dots, y_r, \varepsilon_{r+1}, \dots, \varepsilon_{T-s}, u_{T-s+1}, \dots, u_T)$
- (v) $(v_1, \dots, v_r, \varepsilon_{r+1}, \dots, \varepsilon_{T-s}, y_{T-s+1}, \dots, y_T)$
- (vi) $(v_1, \dots, v_r, \varepsilon_{r+1}, \dots, \varepsilon_{T-s}, u_{T-s+1}, \dots, u_T)$

Proof Proposition

Let \sim denote equivalence in information sets. We prove $(i) \sim (ii)$, $(i) \sim (iii)$, $(ii) \sim (iv)$, $(iii) \sim (v)$ and $(i) \sim (vi)$. This is sufficient to show that all information sets are equivalent.

Case 1: $(i) \sim (ii)$

Note that (ii) $(y_1, \dots, y_r, u_{r+1}, \dots, u_T) = (y_1, \dots, y_r, \phi(L)y_{r+1}, \dots, \phi(L)y_T)$ by using the definition of u in equation (2.6). Since $u_{r+1} = y_{r+1} - \phi_1 y_r - \dots - \phi_r y_1$ with y_1, \dots, y_r and u_{r+1} known, y_{r+1} is known. The same reasoning can be recursively applied to u_{r+2} up to u_T , leading to the desired result.

Case 2: $(i) \sim (iii)$

Note that (iii) $(v_1, \dots, v_{T-s}, y_{T-s+1}, \dots, y_T)$ can be rewritten in terms of only y_t as $(\varphi(L^{-1})y_1, \dots, \varphi(L^{-1})y_{T-s}, y_{T-s+1}, \dots, y_T)$ by using the definition of v_t in equation (2.7). Since $v_{T-s} = y_{T-s} - \varphi_1 y_{T-s+1} - \dots - \varphi_s y_T$ with y_{T-s+1}, \dots, y_T and v_{T-s} known, y_{T-s} is known. The same reasoning can be recursively applied to v_{T-s-1} up to v_1 , leading to the desired result.

Case 3: $(ii) \sim (iv)$

Note that (iv)

$(y_1, \dots, y_r, \varepsilon_{r+1}, \dots, \varepsilon_{T-s}, u_{T-s+1}, \dots, u_T)$ can be rewritten in terms of only y_t as $(y_1, \dots, y_r, \varphi(L^{-1})u_{r+1}, \dots, \varphi(L^{-1})u_{T-s}, u_{T-s+1}, \dots, u_T)$ by using the second equality in equation (2.6). Since $\varepsilon_{T-s} = u_{T-s} - \varphi_1 u_{T-s+1} - \dots - \varphi_s u_T$ with u_{T-s+1}, \dots, u_T and ε_{T-s} known, u_{T-s} is known. The same reasoning can be recursively applied to u_{T-s-1} up to u_{r+1} , leading to the desired result.

Case 4: $(iii) \sim (v)$

Note that (v) $(v_1, \dots, v_r, \varepsilon_{r+1}, \dots, \varepsilon_{T-s}, y_{T-s+1}, \dots, y_T)$ can be represented as $(v_1, \dots, v_r, \phi(L)v_{r+1}, \dots, \phi(L)v_{T-s}, y_{T-s+1}, \dots, y_T)$ by using the second equality in equation (2.7). Since $\varepsilon_{r+1} = v_{r+1} - \phi_1 v_r - \dots - \phi_r v_1$ with v_1, \dots, v_r and ε_{r+1} known, v_{r+1} is known. The same reasoning can be recursively applied to v_{r+2} up to v_{T-s} , leading to the desired result.

Case 5: $(i) \sim (vi)$

To show: $(y_1, \dots, y_T) \sim (v_1, \dots, v_r, \varepsilon_{r+1}, \dots, \varepsilon_{T-s}, u_{T-s+1}, \dots, u_T)$. Denote the vector corresponding to the first information set by \mathbf{y} and the second one by \mathbf{z} . This

statement can be proved using the algebra in Lanne and Saikkonen (2011a). Define the vector $\mathbf{w} = [v_1, \dots, v_{T-s}, u_{T-s+1}, \dots, u_T]'$. Then

$$\begin{bmatrix} v_1 \\ \vdots \\ v_{T-s} \\ u_{T-s+1} \\ \vdots \\ v_T \end{bmatrix} = \begin{bmatrix} y_1 - \varphi_1 y_2 - \dots - \varphi_s y_{s+1} \\ \vdots \\ y_{T-s} - \varphi_1 y_{T-s+1} - \dots - \varphi_s y_T \\ y_{T-s+1} - \phi_1 y_{T-s} - \dots - \phi_r y_{T-s+1-r} \\ \vdots \\ y_T - \phi_1 y_{T-1} - \dots - \phi_r y_{T-r} \end{bmatrix} = \mathbf{A} \begin{bmatrix} y_1 \\ \vdots \\ y_{T-s} \\ y_{T-s+1} \\ \vdots \\ y_T \end{bmatrix}, \quad (2.23)$$

which can be written as $\mathbf{w} = \mathbf{A}\mathbf{y}$. Similarly, we can form the following system of equations

$$\begin{bmatrix} v_1 \\ \vdots \\ v_r \\ \varepsilon_{r+1} \\ \vdots \\ \varepsilon_{T-s} \\ u_{T-s+1} \\ \vdots \\ u_T \end{bmatrix} = \begin{bmatrix} v_1 \\ \vdots \\ v_r \\ v_{r+1} - \phi_1 v_r - \dots - \phi_r v_1 \\ \vdots \\ v_{T-s} - \phi_1 v_{T-s-1} - \dots - \phi_r v_{T-s-r} \\ u_{T-s+1} \\ \vdots \\ u_T \end{bmatrix} = \mathbf{B} \begin{bmatrix} v_1 \\ \vdots \\ v_r \\ v_{r+1} \\ \vdots \\ v_{T-s} \\ u_{T-s+1} \\ \vdots \\ u_T \end{bmatrix}, \quad (2.24)$$

or briefly $\mathbf{z} = \mathbf{B}\mathbf{w}$. Combining both systems of equations, we find that the vectors \mathbf{z} and \mathbf{y} are related in the following way: $\mathbf{z} = \mathbf{B}\mathbf{A}\mathbf{y}$. Since the matrices \mathbf{B} and \mathbf{A} only contain the known causal and noncausal parameters, this shows that these information sets are equivalent. Combining all cases shows that information sets $(i) - (vi)$ are equivalent.

Chapter 3

Do Seasonal Adjustments Induce Noncausal Dynamics in Inflation Rates?

This chapter investigates the effect of seasonal adjustment filters on the identification of mixed causal-noncausal autoregressive models. By means of Monte Carlo simulations, we find that standard seasonal filters induce spurious autoregressive dynamics on white noise series, a phenomenon already documented in the literature. Using a symmetric argument, we show that those filters also generate a spurious noncausal component in the seasonally adjusted series, but preserve (although amplify) the existence of causal and noncausal relationships. This result has important implications for modelling economic time series driven by expectation relationships. We consider inflation data on the G7 countries to illustrate these results.¹

¹This chapter is based on the article Hecq, A., Telg, S. and L. Lieb (2017c), Do Seasonal Adjustments Induce Noncausal Dynamics in Inflation Rates?, *Econometrics*, 5(4), 48, doi:10.3390/econometrics5040048.

3.1 Introduction

Most empirical macroeconomic studies are based on seasonally adjusted data. Various methods have been proposed in the literature aiming at removing unobserved seasonal patterns without affecting other properties of the time series. Just as the misspecification of a trend may cause spurious cycles in detrended data (e.g., Nelson and Kang, 1981), a wrongly specified pattern at the seasonal frequency might have very undesirable effects (see, e.g., Ghysels and Perron, 1993 and Maravall, 1993). A substantial literature focuses on the effects of seasonal adjustment for testing on unit roots, structural change and bias using various autoregressive processes: Ghysels (1990), Ghysels and Perron (1993) and Del Barrio Castro and Osborn (2004) consider the conventional AR, ARMA and periodic AR processes respectively. To the best of our knowledge, there is no result on the consequences of seasonal adjustment for mixed causal-noncausal autoregressive (MAR) process. As discussed in Hecq et al. (2016), MAR models are able to generate features that previously could only be obtained using highly nonlinear and complex models. This includes the possibility of periodic and speculative bubble processes. Following Gouriéroux et al. (2016) and Lanne and Saikkonen (2011a), MAR models can also constitute stationary solutions for rational expectation models.

The present chapter examines the effect of the linear approximation of X-11 seasonal adjustment on the selection of mixed causal-noncausal autoregressive processes. We focus on whether (non)causality in the seasonally adjusted time series is preserved. It is shown that the linear X-11 seasonal adjustment affects the autocovariance function of different MAR specifications of the same autoregressive order in exactly the same way. This is due to the well-known fact that the autocovariance functions and thus spectral densities of these processes are the same (up to a rescaled variance). Since the maximum likelihood procedure used to estimate MAR models exploits more information than solely second-order properties of the process, backward- and forward-looking dynamics are to a great extent preserved in the time series. A simulation study confirms this hypothesis and additionally shows that X-11 seasonal adjustment of white noise series introduces spurious dynamics, which are purely causal or purely noncausal with equal frequency. We investigate the effect of seasonal adjustment on selecting MAR models for inflation rates of the G7 countries Canada, France, Germany, Italy, Japan, the United Kingdom and the United States. It has been found that the distinctive features of inflation can be captured well by MAR models due to their closeness to rational

expectation models (Lanne and Luoto, 2012, 2013; Lanne et al. 2012a). Another reason why noncausal models can provide a good fit to this type of data is the property that they, at least in the Cauchy case, exhibit a causal recursive double autoregressive structure (Gouriéroux and Zakoïan, 2016). This means that one can account for ARCH-type effects in a series by including a noncausal component in the model equation for the conditional mean. For this reason, it is of interest to see how sensitive model selection of inflation is with respect to seasonal adjustment methods.

The remainder of this chapter is organized as follows. Section 3.2 formalizes the notion of mixed causal-noncausal models and comments on the identifiability and estimation of such models. Section 3.3 discusses seasonal adjustment methods based on simple linear filters and mentions their merits and potential pitfalls. The results of the simulation study are collected in Section 3.4. Section 3.5 considers both raw and seasonally adjusted quarterly inflation rates for the G7 countries and shows the effects of seasonal adjustment on the selection of $\text{MAR}(r, s)$ models. Section 3.6 summarizes and concludes.

3.2 Mixed Causal-Noncausal Models

Brockwell and Davis (1991) originally advocated the use of noncausal models as they offer the possibility to rewrite a process with explosive roots in calendar time into a process in reverse time with roots outside the unit circle. Additional important empirical features of the noncausal approach have been put forward in the recent literature. Beyond the improvement in terms of forecasting accuracy (see e.g., Lanne et al., 2012b) as well as their closeness to the concept of nonfundamental shocks (see e.g., Alessi et al., 2011), simple linear noncausal models are able to mimic nonlinear processes such as bubbles or asymmetric cycles (see e.g., Gouriéroux and Zakoïan, 2016; Hecq et al., 2016; Fries and Zakoïan, 2017).

3.2.1 Model Representation

The univariate mixed causal-noncausal autoregressive model $\text{MAR}(r, s)$ for a stationary time series $y_t, (t = 1, \dots, T)$ is written as

$$(1 - \phi_1 L - \dots - \phi_r L^r)(1 - \varphi_1 L^{-1} - \dots - \varphi_s L^{-s})y_t = \varepsilon_t, \quad (3.1)$$

$$\phi(L)\varphi(L^{-1})y_t = \varepsilon_t, \quad (3.2)$$

with L being the backshift operator, i.e., $Ly_t = y_{t-1}$ gives lags and $L^{-1}y_t = y_{t+1}$ produces leads. The error term ε_t is *iid* non-Gaussian. When $\varphi_1 = \dots = \varphi_s = 0$, the process y_t is a purely causal autoregressive process, denoted $\text{AR}(r,0)$ or simply $\text{AR}(r)$:

$$\phi(L)y_t = \varepsilon_t. \quad (3.3)$$

Model specification (3.3) can be seen as the standard backward-looking AR process, with y_t being regressed on y_{t-1} up to y_{t-r} . The process in (3.2) becomes a purely noncausal $\text{AR}(0,s)$ model

$$\varphi(L^{-1})y_t = \varepsilon_t, \quad (3.4)$$

when $\phi_1 = \dots = \phi_r = 0$. Model specification (3.4) is the counterpart of (3.3), since it is a purely forward-looking AR process. That is, y_t does not depend on its past values, but rather on its future values y_{t+1} up to y_{t+s} . Models of the form (3.2) that contain both lags and leads of the dependent variable are called mixed causal-noncausal models.

The roots of both the causal and noncausal polynomials are assumed to lie outside the unit circle, that is $\phi(z) = 0$ and $\varphi(z) = 0$ for $|z| > 1$ respectively. These conditions imply that the series y_t admits a two-sided moving average (MA) representation $y_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}$, such that $\psi_j = 0$ for all $j < 0$ implies a purely causal process y_t (w.r.t. ε_t) and a purely noncausal model when $\psi_j = 0$ for all $j > 0$.

3.2.2 Estimation

The non-Gaussianity assumption ensures the identifiability of the causal and the noncausal part (Breidt et al., 1991). In this chapter, we consider a non-standardized t -distribution for the error process. Lanne and Saikkonen (2011a) show that the parameters of mixed causal-noncausal autoregressive models of the form (3.2) can be consistently estimated by approximate maximum likelihood (AML). Let $(\varepsilon_1, \dots, \varepsilon_T)$ be a sequence of *iid* zero mean t -distributed random variables, then its joint probability density function can be characterized as

$$f_{\varepsilon}(\varepsilon_1, \dots, \varepsilon_T | \sigma, \nu) = \prod_{t=1}^T \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\sigma} \left(1 + \frac{1}{\nu} \left(\frac{\varepsilon_t}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}.$$

The corresponding (approximate) log-likelihood function, conditional on the observed data $\mathbf{y} = [y_1, \dots, y_T]'$ can be formulated as

$$l_y(\boldsymbol{\phi}, \boldsymbol{\varphi}, \alpha, \boldsymbol{\lambda}|\mathbf{y}) = (T - p) \left[\ln \left(\Gamma \left(\frac{\nu + 1}{2} \right) \right) - \ln \left(\sqrt{\nu \pi \sigma^2} \right) - \ln \left(\Gamma \left(\frac{\nu}{2} \right) \right) \right] - \frac{\nu + 1}{2} \sum_{t=r+1}^{T-s} \ln \left(1 + \frac{1}{\nu} \left(\frac{\boldsymbol{\phi}(L)\boldsymbol{\varphi}(L^{-1})y_t - \alpha}{\sigma} \right)^2 \right), \quad (3.5)$$

where $p = r + s$. The distributional parameters are collected in $\boldsymbol{\lambda} = [\sigma, \nu]'$, with σ representing the scale parameter and ν the degrees of freedom. α denotes an intercept that could be introduced in model (3.2), $\Gamma(\cdot)$ denotes the gamma function. Thus, the AML estimator corresponds to the solution

$$\hat{\boldsymbol{\theta}}_{ML} = \arg \max_{\boldsymbol{\theta} \in \Theta} l_y(\boldsymbol{\theta}|\mathbf{y}),$$

with $\boldsymbol{\theta} = [\boldsymbol{\phi}', \boldsymbol{\varphi}', \boldsymbol{\lambda}', \alpha]'$ and Θ is a permissible parameter space containing the true value of $\boldsymbol{\theta}$, say $\boldsymbol{\theta}_0$, as an interior point. Since an analytical solution of the score function is not directly available, gradient based (numerical) procedures (e.g., BHHH and BFGS) can be used to find $\hat{\boldsymbol{\theta}}_{ML}$. If $\nu > 2$, and hence $\mathbb{E}(|\varepsilon_t|^2) < \infty$, the AML estimator is \sqrt{T} -consistent and asymptotically normal. Lanne and Saikkonen (2011a) also show that a consistent estimator of the limiting covariance matrix is obtained from the standardized Hessian of the log-likelihood. For the estimation and selection of mixed causal-noncausal models we follow the procedure proposed in Hecq et al. (2016). Namely, we first estimate by OLS conventional autoregressive models and identify the autoregressive order p that deletes all serial correlation from the residuals. Once this order p has been established, we estimate all combinations of $\text{MAR}(r, s)$ models for which $r + s = p$. The model with the highest log-likelihood in (3.5) at the estimated parameters is selected as final model. It has to be noted that this procedure is not free from misspecification of the noncausal order s for a fixed p as shown in Gouriéroux and Jasiak (2017a).

3.3 Seasonal Adjustment Methods

Seasonal adjustment of data series has received a lot of attention in the econometric and statistical literature; an extensive overview can be found in Bell and Hillmer (2002). The insight that seasonality might alter the legibility of the trend and the cyclical component led to the development of moving averages that adjust series at

their seasonal frequencies. These moving averages are computed using centered, symmetric linear filters of the form

$$\Psi^{SA}(L, L^{-1}) = c_0 + \sum_{j=1}^k c_j (L^j + L^{-j}). \quad (3.6)$$

As these filters are frequently present in seasonal adjustment methods used at statistical agencies (see e.g., Ghysels and Perron, 1993), we focus on them in the simulation study of this chapter.

3.3.1 The Linear X-11 Seasonal Filter

In particular, we focus on the linear approximation of quarterly X-11 seasonal adjustment, $\Psi_{X-11}^{SA}(L, L^{-1})$, which is a moving average of order 57 with weights that sum up to one. The filter has both leads and lags, as it takes 28 quarters before and 28 quarters after every data point into account. This linear approximation represents the result of applying a sequence of symmetric two-sided moving average filters designed to decompose the series into mutually orthogonal trend, seasonal and irregular components (Del Barrio Castro and Osborn, 2004). These operations results in subtracting the estimated seasonal component from the observed series, which gives the linear X-11 seasonal adjustment its representation as time-invariant filtering operation. Table 3.1 shows the final weights, rounded to 3 decimals for expository purposes (see e.g., Ghysels and Perron 1993).

lags/leads		lags/leads		lags/leads	
0	0.856	10	0.025	20	-0.003
1	0.051	11	0.012	21	< 0.001
2	0.041	12	-0.053	22	0.002
3	0.050	13	0.021	23	< 0.001
4	-0.140	14	0.016	24	< 0.001
5	0.055	15	-0.005	25	< 0.001
6	0.034	16	-0.010	26	< 0.001
7	0.029	17	< 0.001	27	< 0.001
8	-0.097	18	0.008	28	< 0.001
9	0.038	19	-0.002		

Table 3.1: Filter weights of the linear quarterly X-11 filter

The Census X-11 program was the most widely applied adjustment procedure by statistical agencies. More recent versions, such as the so-called X-13-ARIMA, consist of first identifying and estimating an ARMA model on the series (with outliers, breaks, calendar effects, etc.) with the aim to extrapolate the variable in the past and in the future before taking a set of moving average filters similar to (3.6). This is done to preserve the number of observations that would be lost in the X-11 method without applying back- and forecasting operations. Many famous statistical agencies and database providers like OECD, NBER and BEA use these methods in order to seasonally adjust most of their series. Additionally, in most econometric software programs (e.g., EViews, RATS, SAS and gretl) various seasonal adjustment approaches based on the X-11 filter are either directly implemented or available through add-on packages that provide these methods. We consider simple linear filters in this study because we want to isolate the effects coming from the moving average adjustment.

3.3.2 Properties of Seasonal Adjustment

The desired property for any seasonal adjustment method is that it only affects the time series of interest at the seasonal frequencies (Ghysels et al., 1993). This requirement is not always fulfilled. For example, suppose we have a zero-mean data series y_t which is seasonally adjusted by a linear filter like in (3.6), with $k = 1$ (for the sake of simplicity). Then the seasonally adjusted series y_t^{SA} has first-order autocovariance equal to:

$$\begin{aligned} \text{Cov}(y_t^{SA}, y_{t-1}^{SA}) &= \text{Cov}((\psi_1 y_{t-1} + \psi_0 y_t + \psi_{-1} y_{t+1}), (\psi_1 y_{t-2} + \psi_0 y_{t-1} + \psi_{-1} y_t)) \\ &= \psi_1 \psi_0 \mathbb{E}(y_{t-1}^2) + \psi_0 \psi_{-1} \mathbb{E}(y_t^2) \\ &= 2\psi_0 \psi_1 \sigma^2, \end{aligned}$$

since $\psi_i = \psi_{-i}$ for all i in the X-11 filter. That is, the seasonally adjusted series y_t^{SA} now has existing autocovariances between observations at t and $t - h$ for $h \neq 0$. Since the autocovariance function is fully symmetric, these autocovariances also exist between observations t and $t + h$ for $h \neq 0$. From convolution theory, it is well-known that the autocorrelation of the linear filter is convolved with the autocorrelation of the series. In other words, the autocorrelation of $\Psi_{X-11}^{SA}(L, L^{-1})$ acts as a smoothing filter on the autocorrelation of y_t . However, if y_t is a white noise (or *iid*) series, these newly existing autocovariances are not due to existing dynamics in the series, but rather spuriously introduced by the seasonal adjust-

ment filter. The special case of applying the X-11 filter to white noise has been well documented in Kaiser and Maravall (2001). We do, however, not rule out this case in our simulation study, as our main interest is to see whether seasonal adjustment filters can create spurious causal and noncausal dynamics in different data generating processes.

3.3.3 Seasonal Adjustment for Mixed Processes

It is common in the seasonality literature to compare the autocovariance function of processes before and after the filtering procedure. In the case of mixed causal-noncausal autoregressions, this procedure does not offer much information. Suppose y_t follows an $\text{MAR}(r, s)$ process as in (3.2) with corresponding autocovariance function $\gamma_y(\cdot)$ and spectrum $f_y(\omega)$. Applying the Filter Theorem, the autocovariances of the X-11 filtered process $y_t^{SA} = \Psi_{X-11}^{SA}(L, L^{-1})y_t$ can be represented as

$$\gamma_{y^{SA}}(k) = \sum_{j=-l}^l \gamma_{\Psi^{SA}}(j) \gamma_y(k-j), \quad (3.7)$$

with $k \in \mathbb{Z}$ and $l = 28$ for the X-11 filter. The sequence $\{\gamma_{\Psi^{SA}}(j) \mid j \in \mathbb{Z}\}$ is given by

$$\gamma_{\Psi^{SA}}(j) = \sum_{i=-\infty}^{\infty} \psi_i \psi_{i+j}, \quad (3.8)$$

with ψ_i being the i th coefficient in the X-11 filter (where $\psi_i = 0$ for $|i| > 28$). However, as it is well-known that a zero-mean $\text{MAR}(r, s)$ process can be represented as a purely causal $\text{AR}(p)$ with exactly the autocovariance function $\gamma_y(\cdot)$ (up to rescaled variance), this measure cannot be used to distinguish between all $\text{MAR}(r, s)$ processes with $r + s = p$. Similarly, due to the fact that also the spectrum $f_y(\omega)$ is the same for all processes of the same total order p , the transfer function $\Phi^{SA}(e^{i\omega})$ and the power transfer function $|\Phi^{SA}(e^{i\omega})|^2$ have exactly the same effect on the second-order properties of e.g. a $\text{MAR}(2,0)$, $\text{MAR}(1,1)$ and $\text{MAR}(0,2)$ process when $p = 2$. Hence, the remaining question is whether (non)causality of the process is preserved in a model selection procedure. To that end, it is interesting to investigate whether the non-Gaussian likelihood procedure can distinguish between different $\text{MAR}(r, s)$ specifications after X-11 seasonal adjustment.

3.4 Simulation Study

In this simulation study, we investigate the effect of X-11 seasonal adjustment on the model selection of various causal, noncausal and mixed causal-noncausal processes.

3.4.1 Purely Causal and Noncausal Processes

We consider three data generating processes for the stationary time series y_t, y_t^- and y_t^+ :

$$y_t = \sum_{s=1}^4 \delta_s D_{s,t} + \varepsilon_t,$$

$$y_t^- = \phi_1 y_{t-1}^- + \sum_{s=1}^4 \delta_s D_{s,t} + \varepsilon_t,$$

$$y_t^+ = \varphi_1 y_{t+1}^+ + \sum_{s=1}^4 \delta_s D_{s,t} + \varepsilon_t,$$

where $D_{s,t}$ ($s = 1, \dots, 4$) are quarterly seasonal dummies with values 1 for the corresponding quarter and zero otherwise. The vector of dummy coefficients is given by $\boldsymbol{\delta} = [-6, 1.5, -0.5, 5]'$ and the autoregressive parameters ϕ_1 and φ_1 are set to 0.7. y_t is a strong white noise augmented with deterministic quarterly seasonality. The processes y_t^- and y_t^+ are a causal and noncausal AR(1) respectively, once again augmented with quarterly dummies. For the three processes the error term ε_t is *iid* t -distributed with 3 degrees of freedom.² The spectra of these processes can be found in Figure 3.1. It can be observed that there are strong peaks at the frequencies $\omega = \frac{1}{2}\pi$ and $\omega = \pi$, which means that there is strong seasonality present in all data generating processes. The computation of these theoretical spectra can be found in Appendix 3.A.

For each series, we consider two types of seasonal adjustment methods. In the first case, we apply the X-11 linear seasonal filter $\Psi_{X-11}^{SA}(L, L^{-1})$ and perform a model selection on the adjusted series denoted $y_t^{SA}, y_t^{-,SA}$ and $y_t^{+,SA}$. In the second case, the original variables y_t, y_t^- and y_t^+ are regressed on four quarterly deterministic seasonal dummies. Model selection is performed afterwards on the

²Taking degrees of freedom equal to 1,2,...,5 give similar qualitative results. As shown in Hecq et al. (2016), identification in finite samples becomes more difficult when the degrees of freedom parameter is high. In practice, a value around 10 might already be considered troublesome.

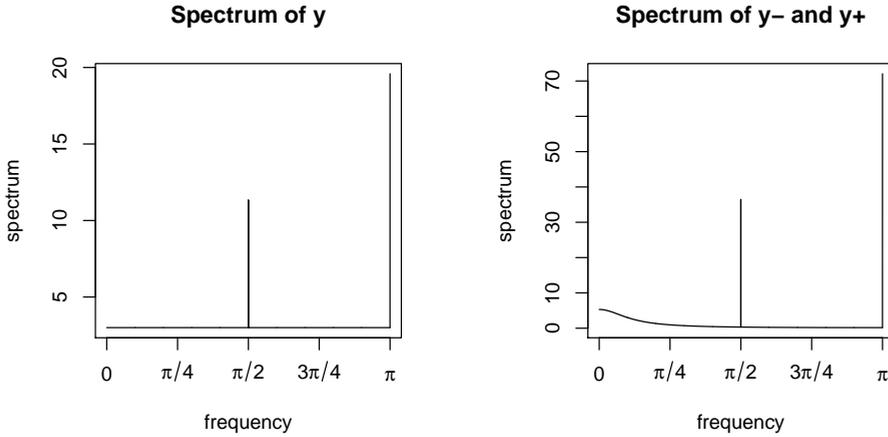


Figure 3.1: Spectra of data generating processes y_t (left), y_t^- and y_t^+ (right)

residuals from this regression, denoted as y_t^D , $y_t^{-,D}$ and $y_t^{+,D}$. In both cases, the model selection is performed as follows. $MAR(r, s)$ models are estimated on the seasonally adjusted series by the AML estimator (assuming a Student's t -distribution) as described in Section 3.2.2, for $r + s = p$ where $p = 1, \dots, 4$ (which accounts for a total of fifteen models). We then rely on BIC for selecting the specification that minimizes that criterion. The results for the first case are collected in Table 3.2 and for the second case in Table 3.4. We display the results for three different sample sizes ($T = 100, 400$ and 700); 1000 replications are used and we add a burn-in period of 50 observations in both sides to delete the possible effect of initial and terminal values on the simulated series.

Case 1: X-11 Seasonally Adjusted Series

Table 3.2 reports the frequencies with which BIC selects the different $MAR(r, s)$ specifications on y_t^{SA} , $y_t^{-,SA}$ and $y_t^{+,SA}$. At $T = 100$, we see that the percentages with which the correct model is selected lie around 80% for all series. These results are not extremely bad, but relatively low when compared to usual results by BIC. Furthermore, it can be seen that the remaining percentages mostly go to either a $MAR(r, s)$ of one order higher than the true data generating process or to models with $r + s = 4$, where especially the purely causal and noncausal specifications are selected. When T increases, the frequency with which a model of order 4 is

selected increases by quite a margin for $y_t^{-,SA}$ and $y_t^{+,SA}$ and drastically for y_t^{SA} . The causal and noncausal AR(1) specifications are still selected with percentages ranging from 72% to 85%. However, these frequencies decrease as T increases (despite $T = 400$ performing better than $T = 100$). For $T = 700$, we see that in 97.3% of the cases BIC either selects a MAR(4,0) or MAR(0,4) for y_t^{SA} instead of the correct white noise specification. (Partial) autocorrelation functions are heavily affected, the X-11 filter creates artificial autocorrelation up to order four due to the large weight at that order in the $\Psi_{X_{11}}^{SA}(L, L^{-1})$ filter. In an almost equal amount of cases, the purely causal MAR(4,0) and purely noncausal MAR(0,4) maximize the log-likelihood (or similarly minimize BIC).

	$T = 100$			$T = 400$			$T = 700$		
	y_t^{SA}	$y_t^{-,SA}$	$y_t^{+,SA}$	y_t^{SA}	$y_t^{-,SA}$	$y_t^{+,SA}$	y_t^{SA}	$y_t^{-,SA}$	$y_t^{+,SA}$
MAR(0,0)	75.5	0.0	0.0	19.5	0.0	0.0	1.9	0.0	0.0
MAR(1,0)	6.4	82.1	4.4	4.3	83.1	0.0	0.3	73.8	0.0
MAR(0,1)	5.1	5.4	82.7	3.8	0.0	84.5	0.5	0.0	72.2
MAR(2,0)	0.6	4.1	0.2	0.2	1.8	0.0	0.0	1.9	0.0
MAR(1,1)	0.5	3.2	3.6	0.0	4.4	4.3	0.0	5.2	5.4
MAR(0,2)	0.2	0.4	3.6	0.4	0.0	1.9	0.0	0.0	1.9
MAR(3,0)	0.6	0.8	0.0	0.0	0.6	0.0	0.0	0.2	0.0
MAR(2,1)	0.1	0.3	0.3	0.0	0.2	1.0	0.0	0.0	1.2
MAR(1,2)	0.0	0.0	0.3	0.0	0.4	0.1	0.0	0.6	0.1
MAR(0,3)	0.1	0.0	0.7	0.0	0.0	0.8	0.0	0.0	0.7
MAR(4,0)	4.8	2.6	0.2	38.3	8.3	0.0	50.6	16.9	0.0
MAR(3,1)	0.0	0.2	0.8	0.2	0.0	0.9	0.0	0.0	1.8
MAR(2,2)	0.7	0.3	0.3	0.0	0.0	0.1	0.0	0.1	0.0
MAR(1,3)	0.3	0.3	0.1	0.0	1.2	0.0	0.0	1.3	0.0
MAR(0,4)	5.1	0.3	2.8	33.3	0.0	6.4	46.7	0.0	16.7

Table 3.2: Frequency (in percentages) with which model is selected (X-11 seasonal adjustment), $\phi_1 = \varphi_1 = 0.7$

The effect of X-11 seasonal adjustment on a simulated white noise series can be graphically observed in Figure 3.2(a). The graph corresponds to the spectrum of the series, which (roughly said) shows how much a certain frequency contributes to the variance of the series. For example, peaks at frequencies $\omega = 0, \pi/2$ and π imply that the series contains a trend component and a seasonal component, associated with the once- and twice-a-year frequencies. This is a well-known pattern in quarterly time series data. As we can see in Figure 3.1(a), the process y_t has a flat spectrum (as for white noise every frequency contributes equally to the

variance of the series) with peaks at the seasonal frequencies. For y_t^{SA} , the peaks at $\omega = 0, \pi/2$ and π have been deleted as desired. However, we observe that not only the seasonal frequencies are affected and the spectrum coincides more with that of a MAR(4,0) model.³ Kaiser and Maravall (2001) comment on this feature by stating that the X-11 procedure is likely to induce spurious cycles in a white noise series.

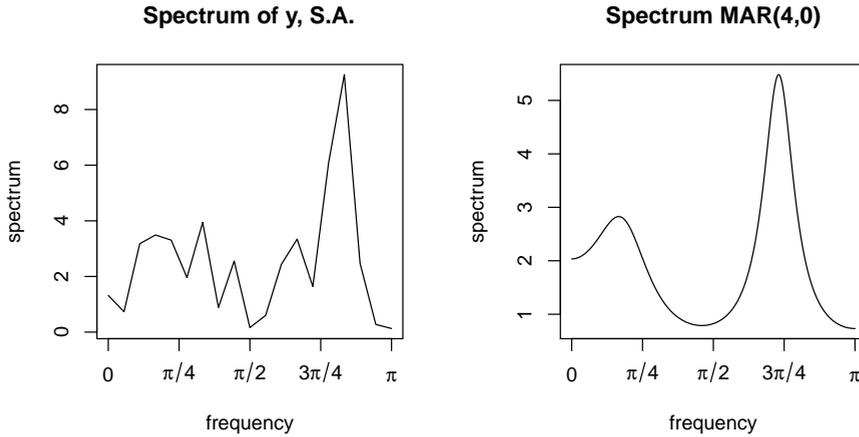


Figure 3.2: Spectra of y_t seasonally adjusted (left) and an MAR(4,0) process (right).

For the first-order autoregressive processes, we see a different pattern arising. Since computing spectra is fully based on the autocovariance generating function, one is unable to distinguish between purely causal and noncausal specifications. This means that the spectra of an MAR(1,0) and MAR(0,1) look exactly the same. Hence, Figure 3.3 shows the spectrum of both the adjusted y_t^- and y_t^+ process. We observe that the X-11 seasonal adjustment filter has affected the spectrum of the series, however only moderately when compared to the white noise case. The time series properties are affected in such a way that higher order models (in particular of order 4) are substantially overselected by BIC, despite the fact that BIC is known for selecting parsimonious models.⁴ Interestingly, information

³The height of the peaks in this theoretical spectrum can be controlled by adjusting the autoregressive parameters. Here we chose $\phi = [0.05, 0.10, 0.25, -0.30]'$ to obtain a spectrum similar to the process y_t^{SA} .

⁴The same simulation exercise has been done based on a seasonal adjustment method called CAMPLET (see Abeln and Jacobs, 2015), which is not based on linear filters. Results show that

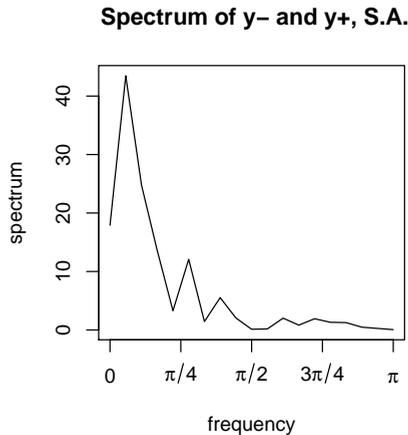


Figure 3.3: Spectra of y_t^- and y_t^+ seasonally adjusted.

in other than first and second order moments seems to remain intact, since BIC selects almost only causal [noncausal] models for the seasonally adjusted causal [noncausal] DGP.

This simulation study has also been performed for different values of ϕ_1 and φ_1 and yielded similar results.⁵ There is, however, one major difference visible. For this reason, we report the results for the case where $\phi_1 = \varphi_1 = 0.3$ in Table 3.3. Note that the data generating process of y_t (i.e., the static case), does not involve the parameters ϕ_1 and φ_1 and hence the results are identical to those in Table 3.2. For the two AR(1) processes, we see that also in this case, (non)causality is mostly preserved. The difference, however, lies in the fact that the higher order models are selected more often and faster in this case. For $\phi_1 = \varphi_1 = 0.7$, first order models were selected in approximately 85 – 90% of the case when $T = 100$ and declined to 70 – 75% when $T = 700$. For $\phi_1 = \varphi_1 = 0.3$, first order models are selected in 80 – 85% when $T = 100$ and decline heavily to 25 – 30% when $T = 700$. This is due to the fact that the autoregressive coefficient is relatively low and thus lies much closer to the coefficients of the X-11 filter (see Table 3.1). Accordingly, the value of the autocovariance at displacement $k = 1$ is also lower (in absolute

MAR(r, s) with $r + s = 4$ are selected in most of the cases. Causality and noncausality is mostly preserved, but not to the same extent as for the X-11. Results are available upon request.

⁵Simulation results are available upon request.

value), which makes it difficult to distinguish it from the autocovariances that are induced by the X-11 filtering process.

	$T = 100$			$T = 400$			$T = 700$		
	y_t^{SA}	$y_t^{-,SA}$	$y_t^{+,SA}$	y_t^{SA}	$y_t^{-,SA}$	$y_t^{+,SA}$	y_t^{SA}	$y_t^{-,SA}$	$y_t^{+,SA}$
MAR(0,0)	75.5	1.6	1.3	19.5	0.0	0.0	1.9	0.0	0.0
MAR(1,0)	6.4	73.1	7.9	4.3	58.1	0.4	0.3	25.4	0.0
MAR(0,1)	5.1	9.1	75.8	3.8	0.1	58.5	0.5	0.1	26.6
MAR(2,0)	0.6	2.5	0.5	0.2	1.2	0.0	0.0	0.7	0.0
MAR(1,1)	0.5	1.7	1.8	0.0	2.4	1.4	0.0	1.1	1.0
MAR(0,2)	0.2	0.2	1.3	0.4	0.0	1.3	0.0	0.0	0.6
MAR(3,0)	0.6	0.3	0.1	0.0	0.1	0.0	0.0	0.0	0.0
MAR(2,1)	0.1	0.3	0.2	0.0	0.1	0.4	0.0	0.2	0.4
MAR(1,2)	0.0	0.0	0.0	0.0	0.2	0.2	0.0	0.2	0.0
MAR(0,3)	0.1	0.0	0.4	0.0	0.0	0.1	0.0	0.0	0.0
MAR(4,0)	4.8	7.6	1.9	38.3	36.2	0.6	50.6	71.2	0.2
MAR(3,1)	0.0	0.1	0.4	0.2	0.0	0.7	0.0	0.0	0.5
MAR(2,2)	0.7	0.6	1.1	0.0	0.0	0.0	0.0	0.0	0.0
MAR(1,3)	0.3	0.4	0.0	0.0	0.9	0.0	0.0	0.9	0.1
MAR(0,4)	5.1	2.5	7.3	33.3	0.7	36.4	46.7	0.2	70.6

Table 3.3: Frequency (in percentages) with which model is selected (X-11 seasonal adjustment), $\phi_1 = \varphi_1 = 0.3$

Case 2: Deterministic Seasonal Adjustment

If we now use seasonal dummy adjusted variables instead of the seasonal filtered ones, we do not see the patterns of Case 1. In Table 3.4, the amount of times the correct model is selected is high. Frequencies increase with the number of time observations, making the selection consistent. In the few cases the right model is not selected, the chosen model has at most a single order more than the correct specification. For y_t^D it selects almost equally the MAR(1,0) and MAR(0,1), while for $y_t^{-,D}$ it picks either the causal MAR(2,0) or the MAR(1,1). For $y_t^{+,D}$, the noncausal MAR(0,2) and MAR(1,1) are often the second best choice.

In the light of cases 1 and 2, we conclude that using raw series and exploiting the correct deterministic seasonal features of the series, does not induce spurious dynamics. X-11 type of filters typically create both causal and noncausal autoregressive parts (due to their two-sidedness) especially if they are applied to white noise series. We do, however, not claim that these filters should not be used. The removal of seasonality remains a challenging task. The data generating pro-

	$T = 100$			$T = 400$			$T = 700$		
	y_t^D	$y_t^{-,D}$	$y_t^{+,D}$	y_t^D	$y_t^{-,D}$	$y_t^{+,D}$	y_t^D	$y_t^{-,D}$	$y_t^{+,D}$
MAR(0,0)	92.9	0.0	0.0	93.9	0.0	0.0	96.6	0.0	0.0
MAR(1,0)	2.5	89.2	2.9	2.7	95.8	0.0	1.3	96.4	0.0
MAR(0,1)	3.2	3.2	90.3	2.9	0.0	96.4	1.9	0.0	97.3
MAR(2,0)	0.3	3.6	0.1	0.2	2.1	0.0	0.1	1.5	0.0
MAR(1,1)	0.4	2.4	2.4	0.2	1.8	1.8	0.0	1.5	0.7
MAR(0,2)	0.2	0.3	2.7	0.1	0.0	1.6	0.1	0.0	1.8
MAR(3,0)	0.1	0.5	0.0	0.0	0.0	0.0	0.0	0.1	0.0
MAR(2,1)	0.1	0.2	0.5	0.0	0.3	0.1	0.0	0.2	0.0
MAR(1,2)	0.1	0.2	0.3	0.0	0.0	0.1	0.0	0.2	0.0
MAR(0,3)	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.2
MAR(4,0)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
MAR(3,1)	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
MAR(2,2)	0.0	0.2	0.1	0.0	0.0	0.0	0.0	0.1	0.0
MAR(1,3)	0.0	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0
MAR(0,4)	0.2	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0

Table 3.4: Frequency (in percentages) with which model is selected (no X-11 adjustment), $\phi_1 = \varphi_1 = 0.7$

cesses considered here only take into account deterministic seasonality. In case of data containing stochastic seasonality, deterministic terms (like e.g., quarterly dummies) will not capture the true seasonal dynamics (see Bell, 1987, for the connection between deterministic and stochastic seasonality).

3.4.2 Mixed Causal-Noncausal Processes

For the purely causal and noncausal processes, we have seen that X-11 seasonal adjustment might heavily affect model selection. The other seasonal adjustment method, where we use an auxiliary dummy regression to remove the seasonality, does not introduce these side effects. This is of course no surprise, as the seasonality introduced in the DGP is indeed deterministic. For this reason, we only consider the effect of X-11 seasonal adjustment for mixed causal-noncausal processes. In particular, we consider the following three data generating processes for the stationary time series y_t^- , y_t^+ and y_t :

$$(1 - \phi_1 L)(1 - \varphi_1 L^{-1})y_t = \sum_{s=1}^4 \delta_s D_{s,t} + \varepsilon_t,$$

$$(1 - a_1L - a_2L^2)y_t^- = \sum_{s=1}^4 \delta_s D_{s,t} + \varepsilon_t,$$

$$(1 - b_1L^{-1} - b_2L^{-2})y_t^+ = \sum_{s=1}^4 \delta_s D_{s,t} + \varepsilon_t,$$

where the first process y_t is a MAR(1,1) process, the second one y_t^- is a causal AR(2) process and the third process y_t^+ is a noncausal AR(2) process. All processes are augmented with seasonal dummies. Once again, we take the vector of dummy coefficients given by $\delta = [-6, 1.5, -0.5, 5]'$ and the error term ε_t to be *iid* t -distributed with 3 degrees of freedom. The autoregressive coefficients are taken as $[a_1, a_2]' = [b_1, b_2]' = [\phi_1, \varphi_1]' = [0.3, 0.4]'$. The results of the model selection procedure for these processes can be found in Table 3.5.

	$T = 100$			$T = 400$			$T = 700$		
	y_t^{SA}	$y_t^{-,SA}$	$y_t^{+,SA}$	y_t^{SA}	$y_t^{-,SA}$	$y_t^{+,SA}$	y_t^{SA}	$y_t^{-,SA}$	$y_t^{+,SA}$
MAR(0,0)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
MAR(1,0)	9.6	2.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0
MAR(0,1)	18.2	0.4	1.9	0.4	0.0	0.0	0.0	0.0	0.0
MAR(2,0)	2.4	61.7	2.8	0.0	24.5	0.0	0.0	5.0	0.0
MAR(1,1)	57.2	5.0	5.0	92.3	0.0	0.0	90.6	0.0	0.0
MAR(0,2)	4.7	2.2	62.0	0.0	0.0	26.9	0.0	0.0	6.5
MAR(3,0)	0.6	2.3	0.2	0.0	0.2	0.0	0.0	0.1	0.0
MAR(2,1)	1.7	2.8	0.4	3.8	1.5	0.0	5.3	0.6	0.0
MAR(1,2)	2.0	0.7	2.5	2.1	0.0	1.5	3.5	0.0	0.8
MAR(0,3)	0.9	0.1	2.6	0.0	0.0	0.4	0.0	0.0	0.1
MAR(4,0)	0.4	16.8	1.4	0.1	72.9	0.1	0.0	94.1	0.0
MAR(3,1)	0.3	1.4	1.4	0.3	0.5	0.4	0.4	0.1	0.0
MAR(2,2)	0.4	1.0	1.2	0.0	0.3	0.6	0.0	0.1	0.3
MAR(1,3)	0.3	2.1	1.1	0.0	0.1	0.5	0.0	0.0	0.2
MAR(0,4)	1.3	1.3	17.3	0.4	0.0	69.6	0.2	0.0	92.1

Table 3.5: Frequency (in percentages) with which model is selected (X-11 seasonal adjustment) for the MAR(1,1), MAR(2,0) and MAR(0,2)

It can be seen that the results for the causal and noncausal AR(2) are relatively similar to the results found for their AR(1) counterparts in the previous section. That is, as T grows larger, the frequency with which BIC selects an autoregressive model of order 4 increases. For the MAR(1,1), this pattern does not arise. For $T = \{400, 700\}$, the MAR(1,1) model is selected in approximately 90% of the cases and thus there is no tendency to select models of order 4. This result seems to

suggest that the overselection of models of higher orders does not play a role for the mixed causal-noncausal model. However, if we allow p to take values higher than 4, we see that also in this case, models of higher dynamic orders are selected.

	$T = 100$	$T = 400$	$T = 700$		$T = 100$	$T = 400$	$T = 700$
MAR(1,0)	8.6	0.0	0.0	MAR(4,2)	1.1	0.5	0.0
MAR(0,1)	11.6	0.0	0.0	MAR(3,3)	0.2	0.0	0.0
MAR(2,0)	1.5	0.0	0.0	MAR(2,4)	1.0	0.3	0.0
MAR(1,1)	36.2	12.2	0.4	MAR(1,5)	4.7	34.5	44.4
MAR(0,2)	3.2	0.0	0.0	MAR(0,6)	1.1	0.2	0.0
MAR(3,0)	0.2	0.0	0.0	MAR(6,1)	0.1	1.7	1.7
MAR(2,1)	0.9	0.1	0.0	MAR(5,2)	0.4	1.8	1.8
MAR(1,2)	3.8	0.7	0.0	MAR(4,3)	0.1	0.3	0.0
MAR(0,3)	1.5	0.0	0.0	MAR(3,4)	0.1	0.1	0.0
MAR(4,0)	0.4	0.0	0.0	MAR(2,5)	0.8	0.5	1.2
MAR(3,1)	0.5	0.0	0.0	MAR(1,6)	1.5	4.4	3.5
MAR(2,2)	0.4	0.0	0.0	MAR(0,7)	0.3	0.0	0.0
MAR(1,3)	0.8	0.0	0.0	MAR(8,0)	0.1	0.0	0.0
MAR(0,4)	0.8	0.0	0.0	MAR(7,1)	0.1	0.0	0.0
MAR(5,0)	1.1	0.0	0.0	MAR(6,2)	0.1	0.0	0.0
MAR(4,1)	3.8	6.6	1.6	MAR(5,3)	0.1	0.1	0.2
MAR(2,3)	0.6	0.1	0.0	MAR(4,4)	0.4	0.0	0.0
MAR(1,4)	5.5	3.2	1.0	MAR(3,5)	0.2	0.1	0.2
MAR(0,5)	1.4	0.2	0.0	MAR(2,6)	0.2	0.0	0.0
MAR(6,0)	0.1	0.0	0.0	MAR(1,7)	1.1	1.4	0.8
MAR(5,1)	2.2	30.6	43.0	MAR(0,8)	1.2	0.4	0.2

Table 3.6: Frequency (in percentages) with which model is selected (X-11 seasonal adjustment) for the MAR(1,1)

The results for the model selection procedure on the X-11 seasonally adjusted MAR(1,1) process y_t when allowing for a maximum autoregressive order of $p = 8$ can be found in Table 3.6.⁶ For $T = 100$, the MAR(1,1) model is selected most of the times, but still only accounts for 36.2% of all cases. The remaining 63.8% is scattered over many alternatives. When T grows, the MAR(1,1) model is selected less often and mostly the MAR(5,1) and MAR(1,5) are selected. For $T = 700$, these two models are selected in 87.4% of the cases. Hence, (non)causality is preserved in the sense that the model selection procedure on a X-11 seasonally adjusted MAR process leads to the selection of models that are also mixed causal-noncausal. This might explain why no models of order 4 were selected in Table 3.5. The X-11 has a high filter weight at displacement 4. However, as the model selection procedure seems to preserve (non)causality, it will also select a model

⁶In order to conserve space, we only report the models that were selected at least once for any $T = \{100, 400, 700\}$.

with at least one lag and one lead. In that case, the only options for $p = 4$ are the MAR(3,1), MAR(2,2) and MAR(1,3), which never reach displacement 4 both in its lags or leads. By allowing p to take values up to 8, there are mixed models that can have lag or leads at higher displacements. Since the coefficient at lag/lead 5 in the X-11 filter is the third highest (not taking into account displacement 0), it creates a nonzero autocovariance in the series y_t^{SA} at order 5. When one estimates an autoregressive model on this adjusted series, a coefficient significantly different from zero in both the causal (lag 5) and the noncausal part (lead 5) is detected. Hence, within $p = 6$, the model with the highest loglikelihood (similarly, the model that minimizes BIC) will be selected. From the results we can see that the MAR(5,1) and MAR(1,5) are being selected in an almost equal amount of the cases. It is important to note that, if we increase p to take values up to 8 for the purely causal and noncausal data generating processes considered in Section 4.1, most frequency is also assigned to the MAR(5,0) and MAR(0,5) respectively.

One may conclude that the seasonal adjustment methods based on linear filters often increase the dynamic order detected by a model selection procedure based on information criteria. However, as we have seen for the case of an MAR(1,1), this order might not be found when the maximum order p on which the model selection procedure is based is not high enough. Hence, a careful inspection of the residuals of the chosen model is highly advised. Note that this feature does not emerge if we carry out a regression with seasonal dummies before investigating the series.

3.5 Seasonal Adjustment, Noncausality and Inflation Rates

A justification for modelling inflation with MAR models is given in Lanne and Luoto (2013). They show that the hybrid New Keynesian Phillips Curve (NKPC) in its regression form,

$$\pi_t = \gamma_f \mathbb{E}_t(\pi_{t+1}) + \gamma_b \pi_{t-1} + \beta x_t + \epsilon_t,$$

where π_t denotes inflation, $\mathbb{E}_t(\cdot)$ the conditional expectation at time t , x_t a measure for marginal costs and ϵ_t an *iid* error term, can be represented as an MAR(r, s) model as in (3.2). More generally, Lanne and Saikkonen (2011a) and Gouriérouc

et al. (2016) discuss how the mixed causal-noncausal model can constitute a stationary solution to linear rational expectation models.

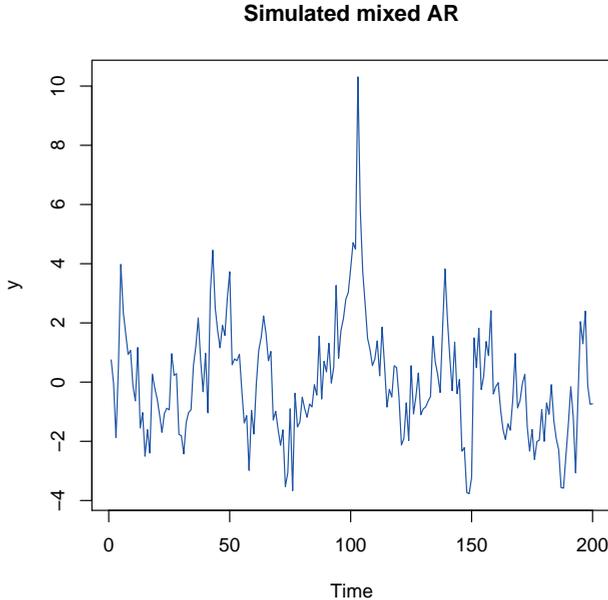


Figure 3.4: Simulated MAR(2,2) process with $\phi = [0.2, 0.3]'$, $\varphi = [0.3, 0.1]'$, $\varepsilon_t \sim t_5$

Figure 3.4 shows a simulated path of an MAR(2,2) with t -distributed error term. It can be seen that such models are able to capture the well-known fluctuating behavior of inflation series. Note that peaks and troughs can be artificially created by considering a more leptokurtic error distribution; choosing the sum of elements in the parameter vector ϕ (respectively φ) close to unity, increases the causal polynomial (respectively noncausal) as driving force in the process.

3.5.1 Data

We consider Consumer Price Index (CPI) series for the Group of Seven (G7) countries Canada, France, Germany, Italy, Japan, United Kingdom and United States. Raw data are available at the OECD database both at the monthly and quarterly frequency. Despite the fact that quarterly data are directly available at OECD, we do not consider those. The reason for this is that those quarterly observations are computed as the unweighted average over the three months in

that quarter. Hence, these data are constructed using a linear filter, which is undesirable in this study as we want to investigate the effect of seasonal adjustment methods based on linear filters. As a consequence, we use quarterly data using point-in-time sampling. That is, we sample the monthly series at the end of each quarter to obtain quarterly data ranging from 1960Q1 until 2017Q2, which accounts for 230 observations. Since prices are stock rather than flow data, this transformation can be considered more natural than averaging.

We examine quarterly observations in order to compare our findings with results found in papers like Lanne and Luoto (2012, 2013) and Lanne et al. (2012a) and in order to make the detection of seasonal unit roots easier. Moreover, X-11 type of filters for monthly data involve a moving average of order 68 and make the identification of the dynamics very difficult as very large displacements have to be considered. Similar to the simulation study, we investigate series in two ways: seasonality is treated (i) deterministically or (ii) by means of seasonal adjustment procedures based on the X-11 filter.

In the first case, we apply a simple seasonal unit root test (HEGY test⁷, see Hylleberg et al., 1990) on the natural logarithm of the raw prices, i.e.,

$$\Delta_4 y_t = \sum_{s=1}^4 \beta_s D_{s,t} + \gamma T_t + \pi_1 z_{1,t-1} + \pi_2 z_{2,t-1} + \pi_3 z_{3,t-2} + \pi_4 z_{3,t-1} + \sum_{i=1}^p \zeta_i \Delta_4 y_{t-i} + \varepsilon_t,$$

where D_{st} are seasonal dummies, T_t is a time trend and $\Delta_4 = (1 - L^4)$, $z_{1,t} = (1 + L + L^2 + L^3)y_t$, $z_{2,t} = -(1 - L + L^2 - L^3)y_t$ and $z_{3,t} = -(1 - L^2)y_t$. Three test-statistics are computed: (i) $H_0 : \pi_1 = 0$: unit root at the zero frequency (nonseasonal stochastic trend), (ii) $H_0 : \pi_2 = 0$, this implies two cycles per year, (iii) $H_0 : \pi_3 = \pi_4 = 0$, the series contains roots i and $-i$ (seasonal unit roots at annual frequencies). The following transformations have to be made in order to remove the seasonal and nonseasonal unit roots in y_t : (i) if we fail to reject the null $\pi_1 = 0$, $(1 - L)$, (ii) if $\pi_2 = 0$, $(1 + L)$ and (iii) if $\pi_3 = \pi_4 = 0$, $(1 + L^2)$. The resulting transformed series is checked for an additional unit root at the zero frequency by the Augmented Dickey Fuller (ADF) test, where the standard regression equation is augmented with quarterly dummies. This additional step allows to determine the degree of integration of the inflation rate, but should be

⁷Alternatively, modified (\mathcal{M}) seasonal unit root tests, see e.g., Del Barrio Castro et al. (2017), could be used. It has been shown that these tests have good finite sample size and power properties. However, as we only apply the seasonal unit root test for illustrative purposes, we restrict ourselves to the original HEGY test.

taken with caution in case of bubbles (see Saikkonen and Sandberg, 2016, for more details on testing for a unit root in the presence of noncausality). Since for all series, the null hypothesis of a unit root is rejected, the transformed series are taken as the measure of inflation that is used for the analysis.

In the second case, the data is immediately adapted by a seasonal adjustment filter. More specifically, we use the X-13 seasonal adjustment procedure on the price series. Afterwards, the conventional ADF test is employed on the log transformed data to check for the presence of a unit root. If this null hypothesis cannot be rejected, the first difference is taken to compute inflation rates. Graphs of these raw and seasonally adjusted series can be found in Appendix 3.B.

After having transformed prices, we apply a model selection procedure using information criteria on both raw and seasonally adjusted inflation series.⁸ Since mixed causal-noncausal AR processes are not identified by Gaussian likelihood, the first step in modelling a time series with a potential forward-looking component is to check for signs of noncausality. We first estimate pseudo-causal AR(p) models by OLS and choose the model order $p = r + s$ with $p_{\max} = 8$ using BIC. Then, diagnostic tests for autocorrelation are performed to see whether additional lags are needed. The null hypothesis of normality is tested on the residuals by means of the Jarque-Bera test.⁹ In case this null cannot be rejected, there is no need to consider mixed causal-noncausal models, as the backward- and forward-looking components cannot be distinguished from each other. In case the null of normality is rejected, all MAR(r, s) specification for the selected pseudo-causal order p are considered. As discussed in Section 3.2.2, we assume a non-standardized t -distribution for the error process. The model that maximizes the corresponding log-likelihood is chosen to be the final model.

3.5.2 Results

Table 3.7 shows the results of the HEGY test on the natural logarithm of raw quarterly CPI series. Critical values are from Franses and Hobijn (1997). Rejections of the null hypothesis at a 5% significance level are indicated by asterisks. Test results indicate that the presence of a zero frequency unit root cannot be rejected for all seven countries, while the presence of seasonal unit roots at annual

⁸The \mathcal{R} package MARX used in this study is freely available at <https://CRAN.R-project.org/package=MARX>. See Hecq et al. (2017b) for instructions on how to use the package.

⁹Since the Jarque-Bera test is based on the sample skewness and kurtosis, which might not exist for fat-tailed processes, we also performed the Kolmogorov-Smirnov and Anderson-Darling test to check for normality. These tests confirmed the reported results for the Jarque-Bera test.

frequencies is rejected in all cases. The possibility of prices containing two cycles per year is rejected for all countries. Hence, $(1 - L)$ is the correct transformation to compute inflation for all series. Subsequently, we determine the pseudo-causal model order p , where we include an intercept and quarterly dummies in the regression equation. We find that BIC selects models of order 4 for all countries, except for France and Italy for which a dynamic order of 3 is detected. In all cases, normality of the residuals is rejected using the Jarque-Bera test, which justifies the use of mixed causal-noncausal models. Additionally, all residuals are tested for displaying ARCH effects using Engle's ARCH[1-2] test. This LM test with the null of no-ARCH is rejected for France, Italy, Japan and the United Kingdom. We discuss this feature and how to deal with it in more detail in Section 2.5.3. For the chosen p in the pseudo-causal model, all combinations of models with $p = r + s$ were estimated and the model that maximizes the log-likelihood is shown in the table. For three countries, a purely causal model was selected, while for two countries a purely noncausal model was chosen as final model. The remaining two countries were found to have mixed causal-noncausal dynamics. The estimated degrees of freedom parameter for the countries Italy, Japan and United Kingdom takes values 1.87, 2.14 and 2.39 respectively. For the United States and Canada, the values lie slightly higher with 3.47 and 5.79. As shown in Hecq et al. (2016), the estimates are more precise (in finite samples) when the error distribution exhibits fat tails. Hence, the estimation and corresponding identification (through the value of the log-likelihood at the estimated parameter values) is expected to be relatively precise in these cases. For France and Germany, this cannot be argued, as the estimated degrees of freedom parameter takes values 20.22 and 20.40 respectively. Since this distribution approaches the normal distribution, the identification of mixed causal-noncausal models might be problematic (despite the rejection of normally distributed residuals of the pseudo-causal model).

Country	$H_0 : \pi_1 = 0$	$H_0 : \pi_2 = 0$	$H_0 : \pi_3 = \pi_4 = 0$	Pseudo model	Jarque-Bera: H_0 : normality	ARCH-LM: H_0 : no-ARCH	MAR(r, s)
Canada	-0.91	-6.39*	36.03*	AR(4)	9.99*	1.35	MAR(0,4)
France	-1.51	-5.80*	50.62*	AR(3)	40.26*	20.42*	MAR(3,0)
Germany	-1.05	-5.85*	27.47*	AR(4)	6.33*	1.77	MAR(0,4)
Italy	-0.86	-9.91*	142.52*	AR(3)	292.63*	66.78*	MAR(3,0)
Japan	-1.29	-3.66*	25.94*	AR(4)	121.72*	10.46*	MAR(4,0)
United Kingdom	-1.06	-6.21*	43.70*	AR(4)	191.87*	17.06*	MAR(1,3)
United States	-0.86	-6.60*	33.44*	AR(4)	816.86*	0.82	MAR(2,2)
c.v. (5%)	-3.49	-2.91	-6.57		5.99	3.00	

Table 3.7: HEGY test on prices and the identification of MAR(r, s) models on quarterly inflation rates (not s.a.)

The results for X-13 seasonally adjusted data can be found in Table 3.8.¹⁰ Note that we do not apply the HEGY test to the seasonally adjusted series, as seasonal effects (and thus also seasonal roots) are assumed to be removed by applying the filters. Hence, it suffices to perform the ADF-test for the log transformed price series. Since all series are found to contain a unit root, we take the first difference $(1 - L)$ to compute the growth rate of prices, i.e. inflation. On the transformed series, the pseudo-causal model that adequately captures the serial correlation, is identified. We see that, except for Italy, all countries have an autoregressive order equal to or lower than the one found on raw data. Hence, this is not in line with the results in the simulation study. In all cases, we have the same conclusion for the Jarque-Bera and ARCH-LM test at the 5% significance level when compared to the raw data. That is, the null of normality is rejected for all countries, while the null of no-ARCH effects cannot be rejected for Canada, Germany and the United States. As a result, we can investigate the presence of noncausality. We find that in four out of seven cases, (non)causality is exactly preserved. In these cases, a purely causal [noncausal] model for the raw data series also results in a purely causal [noncausal] model for the seasonally adjusted series. In the other three cases, a mixed causal-noncausal model either switched to a purely causal or noncausal model, or vice versa.

Country	ADF-statistic H_0 : unit root	Pseudo model	Jarque-Bera: H_0 : normality	ARCH-LM: H_0 : no-ARCH	MAR(r, s)
Canada	-0.60	AR(3)	28.11*	2.39	MAR(0,3)
France	-0.86	AR(3)	166.52*	14.21*	MAR(3,0)
Germany	-0.75	AR(3)	51.33*	0.22	MAR(0,3)
Italy	-1.58	AR(6)	754.96*	45.54*	MAR(2,4)
Japan	-1.55	AR(3)	76.95*	23.30*	MAR(3,0)
United Kingdom	-0.59	AR(2)	538.86*	13.05*	MAR(0,2)
United States	-0.54	AR(3)	1320.34*	0.64	MAR(0,3)
c.v. (5%)	-3.43		5.99	3.00	

Table 3.8: Identification of MAR(r, s) models on quarterly inflation rates (s.a)

¹⁰The same investigation has been done using the TRAMO/SEATS seasonal adjustment method (see Maravall, 1997) which is merely used by Eurostat. This method is based on unobserved components decomposition but is not free from filters. In particular, a truncated version of the two-sided, centered, symmetric Wiener-Kolmogorov filter is used to estimate the signal in an observed process y_t (for more details, see e.g., Maravall, 2006). As the results are very similar, we do not report them here.

3.5.3 Considerations

A reason for the different results in the simulation study and the empirical study might be due to the model selection procedure. Gouriéroux and Jasiak (2017a) show that the two-step approach, in which the first step consists of estimating purely causal autoregressive processes by OLS, might lead to an estimated model that is misspecified with respect to its noncausal order. A potential other method (henceforth referred to as one-step method) might be to estimate all models directly by means of the method of maximum likelihood, to compute its log-likelihoods and compute information criteria accordingly. However, Gouriéroux and Jasiak (2017a) show that this method might also fail to reveal the true causal and noncausal orders. It is important to note that the consequences of an order misspecification in mixed causal-noncausal processes are different from the consequences of order misspecification in Gaussian $AR(p)$ processes. Since Gaussian $AR(p)$ models are nested, a potential wrongly estimated autoregressive order has no effect on the consistency of the autoregressive coefficient estimators. In contrast, mixed causal-noncausal processes with $p = r + s$ with fixed p and varying s are non-nested, and the parameter estimates of the autoregressive coefficients resulting from a misspecified model may be inconsistent. We want to emphasize that the non-nestedness of mixed causal-noncausal processes heavily complicates the model selection procedure. Identification in small samples is achieved faster and more accurate when the processes have extreme fat tails (Hecq et al., 2016). However, as these data have highly nonlinear patterns, we observe that adjusting the sample size by a small number of observations might already have big effects. As the relationships between the causal and noncausal parameters are highly nonlinear and complex, we have that a slight deviation in (one of) these values might lead to a drastic effect in the maximization of the log-likelihood in finite samples. An additional complication that arises is that optimization routines are sometimes prone to numerical inaccuracies, which amplifies this concern, since they have a direct impact on the final model selected by information criteria.¹¹

¹¹We performed the one-step method on the raw and seasonally adjusted inflation rates. In many cases, the one- and two-step procedure select the same model. However, the one-step procedure is more sensitive to numerical inaccuracies in the optimization routine, as the number of different models to be estimated heavily increases even when p goes up from e.g. 4 to 8 (from 15 to 45 models). The model selected in the two-step approach is, however, often among the second or third best when ranked by the values of the information criteria. An advantage of the two-step approach is that p is bounded and numerical inaccuracies only play a role for all models estimated within this p .

Another important consideration is concerned with the type of data considered in this application and its effect on model selection. Inflation rates are relatively volatile and might display conditional heteroskedasticity. Hecq (1996) shows that these effects might directly affect the autoregressive order determined by information criteria. As Gouriéroux and Zakoian (2016) and Fries and Zakoian (2017) show that a noncausal Cauchy process can exhibit ARCH-type effects because of its causal recursive double autoregressive structure, it is advisable to check the residuals of the pseudo-causal model for the presence of ARCH effects in the spirit of misspecification analysis. Cavaliere et al. (2017) propose to use an alternative to the standard Lagrange multiplier test (Engle's ARCH LM test) as it is based on conventional asymptotic Gaussian p -values, which is not appropriate in this framework. To circumvent this problem, they suggest a test based on Spearman's rank statistic, which tests non-parametrically the relationship between the residuals and the square of the lagged residuals. Applying this test to the seasonally adjusted data, we fail to reject the null hypothesis of no-ARCH for all series. The test results are however inconclusive, as applying the same test to the residuals of the final chosen $MAR(r, s)$ model lead to a non-rejection for all countries, except the United Kingdom. One reason for this unexpected result might be that the (potential) ARCH dependence present in the data is not of double autoregressive nature. Another reason might be the fact that we are not dealing with noncausal Cauchy processes (all estimated degrees of freedom parameters were strictly higher than 1) and thus the results found for these processes do not extend to our data. In any case, it seems apparent that the development of more robust strategies for testing misspecification and to perform model selection is needed for mixed causal-noncausal models.

As a last remark, it has to be mentioned that the findings in this chapter are likely to extend to cases beyond the exercise of seasonal adjustment. For instance, De Jong and Sakarya (2016) derive a new representation of the Hodrick-Prescott (HP) filter which highlights that it is a symmetric weighted average similar to the filters considered in this chapter. They further state that, in case of a unit root process, the weak dependence of the cyclical component suggests that the unit root is absorbed into the trend component. As the filter (and autocorrelation function) is symmetric, this introduces spurious autocorrelation identically in calendar and reverse time.

3.6 Conclusion

We investigate the effect of seasonal adjustment on model selection for the inflation rate of the G7 countries. In particular, we study whether seasonal adjustment may spuriously affect the noncausality found in different time series. Since raw data are directly available, we can compare model selection by BIC where one (*i*) deterministically removes seasonality or (*ii*) applies a predefined seasonal adjustment filter. We find that the linear X-11 seasonal adjustment procedure induces spurious dynamics in white noise processes, which are purely causal or purely non-causal with equal frequency. In case the data generating process is purely causal or noncausal, we find that the direction of causality is preserved after seasonal adjustment. That is, if a process is noncausal the probability to find a purely causal AR model is relatively small. However, we do find that seasonal adjustment decreases the frequency with which we detect dynamics in favor of a less parsimonious model. For both pure processes (causal and noncausal) and mixed causal-noncausal processes, we find that the dynamic order selected after seasonal adjustment is higher than the true autoregressive order of the process.

In the empirical study, we observe that preservation of (non)causality is indeed confirmed in most cases. It is explained how discrepancies might arise due to the presence of ARCH effects and the potential flaws in the model selection procedure due to non-nestedness of mixed causal-noncausal processes. It is, however, very important to carefully consider seasonally adjusted data as it might lead to misleading conclusions. In terms of the hybrid NKPC, identified forward- and backward-looking behaviour might rather be an artefact from the filtering procedure, than a characteristic of the raw data.

3.A Autocovariances and Spectra

Similar to Ghysels and Perron (1993) and Del Barrio Castro and Osborn (2004), one can obtain the autocovariances and spectra of the processes considered in the simulation study.

White Noise with Seasonal Dummies

The first data generating process (y_t) is given by

$$y_t = \sum_{s=1}^4 \delta_s D_{s,t} + \varepsilon_t, \quad (3.9)$$

i.e. a deterministic (seasonal dummies) and a stochastic (strong white noise) part. This specification can also be written as a model of hidden periodicities (see e.g. Wei, 2006):

$$y_t = \sum_{k=1}^2 \left[a_k \cos\left(\frac{1}{2}\pi kt\right) + b_k \sin\left(\frac{1}{2}\pi kt\right) \right] + \varepsilon_t, \quad (3.10)$$

in which the quarterly seasonal dummies are represented in terms of sines and cosines. The coefficients can be computed with $a_1 = \frac{1}{2} \sum_{s=1}^4 \delta_s \cos\left(\frac{1}{2}\pi s\right)$, $a_2 = \frac{1}{4} \sum_{s=1}^4 \delta_s \cos(\pi s)$ and $b_1 = \frac{1}{2} \sum_{s=1}^4 \delta_s \sin\left(\frac{1}{2}\pi s\right)$. Note that we do not compute the explicit form of b_2 as it contains the term $\sin(\pi s)$ which equals zero for all s . It can be derived that the coefficients in (3.9) and (3.10) are linked as follows: $\delta_1 = b_1 - a_2$, $\delta_2 = -a_1 + a_2$, $\delta_3 = -b_1 - a_2$ and $\delta_4 = a_1 + a_2$. Using these relationships, we can compute the power spectrum:

$$\text{Power} = \frac{1}{2} (a_1^2 + b_1^2) + a_2^2 = \sum_{k=-1}^2 |c_k|^2,$$

where $c_k = \frac{1}{4} \sum_{s=1}^4 \delta_s e^{-i\omega_k s}$ and $\omega_k = \frac{1}{2}\pi k$. With the values for δ_s ($s = 1, \dots, 4$) used in the simulation study, i.e., $\boldsymbol{\delta} = [\delta_1, \delta_2, \delta_3, \delta_4]' = [-6, 1.5, -0.5, 5]'$, we obtain $[a_1, a_2, b_1]' = [1.75, 3.25, -2.75]'$ and $[c_{-1}, c_0, c_1, c_2]' = [0.875 - 1.375i, 0, 0.875 + 1.375i, 3.25]'$ with total power equal to $15\frac{7}{8}$. Following Wei (2006), the power

spectrum $f(\cdot)$ can be attributed to the several frequencies:

$$f = \begin{cases} \frac{1}{2}(1.75^2 + (-2.75)^2) = 5\frac{5}{16}, & \text{for } k = 1, \\ 3.25^2 = 10\frac{9}{16}, & \text{for } k = 2, \\ 0, & \text{otherwise.} \end{cases}$$

When $k = 1$, $\omega_1 = \frac{1}{2}\pi$, $f(\omega_1) = 5\frac{5}{16}/2\pi$ and when $k = 2$, $\omega_2 = \pi$, $f(\omega_2) = 10\frac{9}{16}/2\pi$. Hence, the spectrum of the deterministic part shows spikes at the frequencies $\frac{1}{2}\pi$ and π , but equals zero otherwise. The process y_t in (3.9) also contains a stochastic part (ε_t). It is well-known that the spectrum of a white noise sequence is a flat line (all frequencies contribute equally) equal to $\sigma_\varepsilon^2/2\pi$. Since ε_t is *iid* distributed with 3 degrees of freedom, $\sigma_\varepsilon^2 = \frac{3}{3-2} = 3$. Since the deterministic and stochastic parts are independent processes, the spectrum of y_t can be computed as the sum of the two independent spectra (Wei, 2006). Hence, the spectrum of y_t equals $\frac{3}{2}\pi$ at all frequencies, except for $\omega = \frac{1}{2}\pi$ and $\omega = \pi$. At these frequencies, the power spectrum equals $\frac{133}{32}\pi$ and $\frac{217}{32}\pi$ respectively. The autocovariance structure of the process can also be deduced. Using results from Hamilton (1994, p. 156), we find that $\mathbb{E}(y_t) = 0$, $\mathbb{E}(y_t^2) = \sum_{k=1}^2 \sigma_k^2$ and $\mathbb{E}(y_t y_{t-h}) = \sum_{k=1}^2 \sigma_k^2 \cos(\frac{1}{2}\pi kh)$, where σ_k^2 corresponds to variances $\mathbb{E}(a_k^2)$ and $\mathbb{E}(b_k^2)$ of the coefficient sequences in model (3.10).

AR(1) with Seasonal Dummies

The data generating processes for y_t^- and y_t^+ , are a combination of an AR(1) process and the seasonal dummies. Let us focus on the causal AR(1) representation,

$$y_t^- = \phi y_{t-1}^- + \sum_{s=1}^4 \delta_s D_{s,t} + \varepsilon_t, \quad (3.11)$$

which has a corresponding MA(∞) representation given by

$$\begin{aligned} y_t^- &= \frac{1}{1-\phi} \sum_{s=1}^4 \delta_s D_{s,t} + \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}, \\ &= X_t + Z_t. \end{aligned} \quad (3.12)$$

In order to compute the spectrum of the AR(1) part, we can compute the autocovariance generating function $\gamma(L) = [(1 - \phi L)(1 - \phi L^{-1})]^{-1} \sigma_\varepsilon^2$. From this it follows directly (see e.g. Wei, 2006) that the spectrum equals $f(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{1}{1 + \phi^2 - \phi(e^{-i\omega} + e^{i\omega})} = \frac{\sigma_\varepsilon^2}{2\pi} \frac{1}{1 + \phi^2 - 2\phi \cos(\omega)}$. The autocovariance generating function (and thus spectral density) of both the causal AR(1) and noncausal AR(1) are identical. The spectra of y_t^- and y_t^+ are thus computed as the sum of the spectrum of the autoregressive model (Z_t) and the spectrum of the rescaled dummy model (X_t) discussed above. The peaks at the frequencies $\omega = \frac{1}{2}\pi$ and $\omega = \pi$ are simply rescaled by $\frac{1}{1-\phi}$, as can be seen from (3.12). Figure 3.1 shows that there is indeed strong seasonality present in all data generating processes. The autocovariance function of the processes can also be readily deduced by noting that this process is the sum of two independent processes. Then we obtain $\mathbb{E}(y_t^-) = 0$, $\mathbb{E}[(y_t^-)^2] = \left(\frac{1}{1-\phi}\right)^2 \sum_{k=1}^2 \sigma_k^2 + \frac{\sigma_\varepsilon^2}{1-\phi^2}$ and $\mathbb{E}(y_t^- y_{t-h}^-) = \left(\frac{1}{1-\phi}\right)^2 \sum_{k=1}^2 \sigma_k^2 \cos\left(\frac{1}{2}\pi kh\right) + \frac{\phi^h \sigma_\varepsilon^2}{1-\phi^2}$. The autocovariance generating function and spectra of the purely causal and noncausal AR(2) and MAR(1,1) with seasonal dummies can be computed analogously.

3.B Graphs

This appendix contains the graphs of the inflation rate of the G7 countries used in this chapter. Figure 3.5 contains the rates of inflation computed according to the raw prices series, while Figure 3.6 is computed using X-13 seasonally adjusted price series.

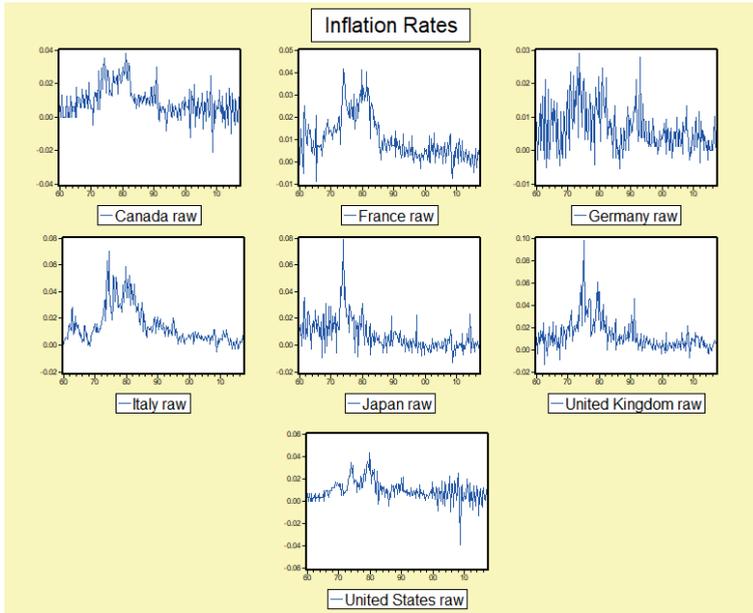


Figure 3.5: Raw inflation rates

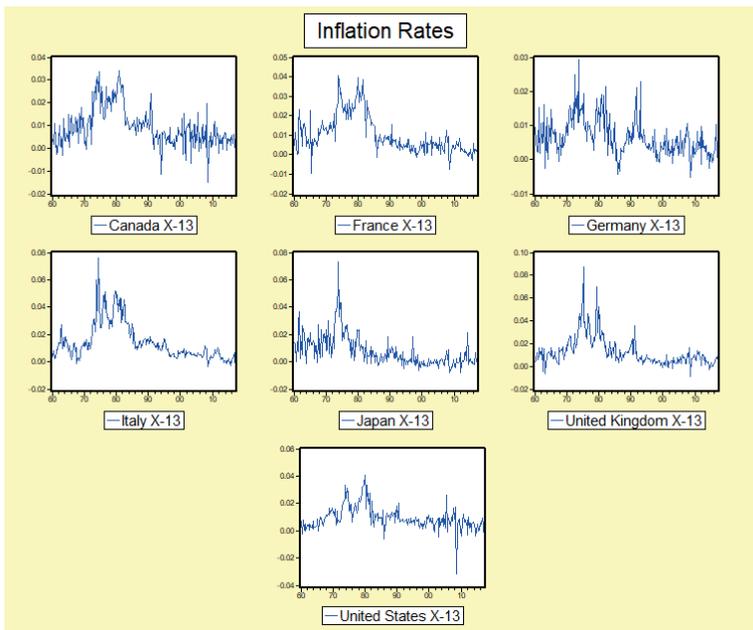


Figure 3.6: X-13 seasonally adjusted inflation rates

Chapter 4

Mixed Causal-Noncausal Autoregressions with Exogenous Regressors

The mixed autoregressive causal-noncausal model (MAR) has been proposed to estimate economic relationships involving explosive roots in their autoregressive part, as they have stationary forward solutions. In previous work, possible exogenous variables in economic relationships are substituted into the error term to ensure the univariate MAR structure of the variable of interest. To allow for the impact of exogenous fundamental variables directly, we instead consider a MARX representation which allows for the inclusion of exogenous regressors. We develop the asymptotic distribution of the MARX parameters. We assume a Student's t -likelihood to derive closed form solutions of the corresponding standard errors. By means of Monte Carlo simulations, we evaluate the accuracy of MARX model selection based on information criteria. We investigate the influence of the U.S. exchange rate and the U.S. industrial production index on several commodity prices.¹

¹This chapter is based on the paper Hecq, A., Issler, J. and S. Telg (2017a), Mixed Causal-Noncausal Autoregressions with Strictly Exogenous Regressors, *MPRA Paper 80767*, University Library of Munich, Germany.

4.1 Introduction

The usefulness of mixed causal-noncausal autoregressive (MAR) models in time series econometrics can be explained by three reasons. First, Gouriéroux and Zakoïan (2016), Hencic and Gouriéroux (2014) and Hecq et al. (2016) demonstrate how the inclusion of noncausal autoregressive terms can generate dynamic patterns like speculative bubbles and asymmetric cycles that instead should be generated using complex nonlinear models. Second, Lanne et al. (2012a; 2012b) have shown that allowing for noncausality might improve forecast performances. Third, the MAR representation of an economic variable can be interpreted as a solution of a rational expectation model (Lanne and Saikkonen, 2011a). Whereas causal autoregressive models take only the fundamental solution of a system into account, the mixed causal-noncausal autoregressive model explicitly allows for non-fundamental outcomes, as the process of interest might depend on past, current and future (nonfundamental) shocks. This extension proves extremely useful, as it is well known that some economic models do not possess a fundamental solution by construction (Alessi et al., 2011) and explicitly consider that the information available to economic agents is larger than the one of econometricians.

The inclusion of exogenous regressors might further improve the relevance of the MAR model for modelling and forecasting economic processes. This model, enriched with exogenous variables, is denoted MARX. In this chapter, we show how the parameters of the MARX model can be estimated by maximum likelihood (ML) for noise driven by a general class of non-Gaussian densities as defined in Andrews et al. (2006). For the particular case of a Student's t -likelihood, we provide a method to compute closed form solutions of the corresponding standard errors. This is also done for the Least Absolute Deviation (LAD) estimator, as it is often used as an initial estimator (Lanne and Saikkonen, 2011a) and is known to outperform the Student's t ML estimator in certain instances (Hecq et al., 2016). We show that purely causal and noncausal models with additional regressors have the appealing feature to be (potentially) identifiable under Gaussianity, due to cross-covariances creating an asymmetry in the autocovariance function of both models. This property is not shared by purely causal and noncausal models without exogenous regressors. By means of Monte Carlo simulations, we evaluate the performance of the ML estimator as well as our proposed model selection method based on information criteria. The methods proposed in this chapter are

implemented in the R package **MARX**; its specifications are discussed in detail in Chapter 6.²

A simple underlying economic example where the MARX model can be used is the New Keynesian Phillips Curve (NKPC) where the solution of the model for the inflation rate depends on the output gap (see e.g., Pesaran, 2015, p. 475). More generally, many relationships that involve expectation terms and consequently have a present value solution, fall in this framework. In the empirical section we investigate the relationship between several commodity prices indices representative of the global market ($cp_t \equiv \ln CP_t$) and two fundamental explanatory variables: the U.S. nominal exchange rate ($s_t \equiv \ln S_t$) and the U.S. industrial production index ($ip_t \equiv \ln IPI_t$) (see Chen et al., 2010; Bork et al., 2014).³ Let us consider a rational expectation (RE) model such as

$$\Delta cp_t = \beta_b \Delta cp_{t-1} + \beta_f \mathbb{E}(\Delta cp_{t+1} | \Omega_t) + \vartheta \Delta s_t + u_t, \quad (4.1)$$

where $\mathbb{E}(\Delta cp_{t+1} | \Omega_t)$ is the expectation made at time t of the future endogenous variable conditional on Ω_t , the information set available at time t .⁴ Using the quadratic determinantal equation method, we can write (4.1) as

$$\Delta cp_t - \alpha_b \Delta cp_{t-1} = \left(\frac{\vartheta}{1 - \beta_f \alpha_b} \right) \sum_{j=0}^{\infty} \alpha_f^{-j} \mathbb{E}(\Delta s_{t+j} | \Omega_t) + \left(\frac{1}{1 - \beta_f \alpha_b} \right) u_t, \quad (4.2)$$

where α_b is one of the two roots of the quadratic equation $\beta_f x^2 - x + \beta_b = 0$ (for more details see Pesaran, 2015, p. 473-475). That is, the RE model can be represented as a lag-augmented present value model where commodity prices Granger-cause exchange rates: a model compatible with our MARX representation.

The remainder of the chapter is organized as follows. Section 4.2 formalizes the notion of MARX models, their identifiability and how to simulate such processes. Section 4.3 considers (approximate) ML estimation and introduces a convenient way to compute standard errors which is not based on computing the Hessian

²The package is freely available at <https://CRAN.R-project.org/package=MARX>.

³Figure 4.2 of the empirical section displays the series we have used for our empirical investigation. The five monthly commodity price indices are the IMF primary commodity price indices and the U.S. exchange rate is released by the Federal Reserve Bank of St. Louis. We also make use of the industrial production index from the FRED database.

⁴We can substitute the expectation term $\mathbb{E}(\Delta cp_{t+1} | \Omega_t)$ either by a perfect foresight scheme $\mathbb{E}(\Delta cp_{t+1} | \Omega_t) = \Delta cp_{t+1}$ or by the sum of their realizations plus the realization of a martingale difference process $\mathbb{E}(\Delta cp_{t+1} | \Omega_t) = \Delta cp_{t+1} + \xi_{t+1}$ (see Broze et al., 1995).

using gradient based numerical procedures. The results from various Monte Carlo simulations are collected in Section 4.4. Section 4.5 details the empirical applications. Section 4.6 summarizes and concludes. Proofs and additional material are collected in the Appendix.

4.2 The MARX Model

Let y_t be the variable of interest which is observed over the time period $t = 1, \dots, T$. Let $x_{i,t}$ ($i = 1, \dots, q$) be the i th variable in a set of q for y_t and $\beta \in \mathbb{R}^q$ a vector of parameters. Then we can define $\mathbf{X}_t = [x_{1,t}, \dots, x_{q,t}]' \in \mathbb{R}^q$ as the vector of all exogenous variables at time t .⁵ The MARX(r, s, q) for a stationary time series y_t can be represented as

$$\phi(L)\varphi(L^{-1})y_t - \beta' \mathbf{X}_t = \varepsilon_t, \quad (4.3)$$

where $\phi(L)$ is a lag polynomial of order r , $\varphi(L^{-1})$ a lead polynomial of order s and $r + s = p$. The operator L is the lag operator when raised to positive powers, i.e., $L^i y_t = y_{t-i}$, and interpreted as a lead operator when raised to negative powers: $L^{-i} y_t = y_{t+i}$. The error term ε_t is assumed to be strong white noise. When $\varphi_1 = \dots = \varphi_s = 0$, the process y_t is a purely causal autoregressive model with exogenous regressors, denoted MARX($r, 0, q$) or simply ARX(r, q):

$$\phi(L)y_t - \beta' \mathbf{X}_t = \varepsilon_t. \quad (4.4)$$

Specification (4.4) can be seen as the standard backward-looking ARX model. Conversely, the process in (4.3) reduces to a purely noncausal MARX($0, s, q$):

$$\varphi(L^{-1})y_t - \beta' \mathbf{X}_t = \varepsilon_t, \quad (4.5)$$

when $\phi_1 = \dots = \phi_r = 0$. Note that the concepts of causality and noncausality are defined in terms of the strictly stationary solution of the model. To that end, we assume that both polynomials in (4.3) have their zeros outside the unit circle:

$$\phi(z) \neq 0 \text{ for } |z| \leq 1 \text{ and } \varphi(z) \neq 0 \text{ for } |z| \leq 1. \quad (4.6)$$

⁵We only consider contemporaneous values of \mathbf{X}_t in the model. The MARX model can also take the form of a mixed autoregressive distributed lag (MARDL) model. See Appendix 4.A for derivation and motivation.

When $q = 0$, the process in (4.4) [(4.5)] reduces to a purely causal [noncausal] AR process that has a one-sided MA representation consisting of only past [future] and current values of ε_t . For the general process in (4.3) these conditions however imply that the process y_t follows a two-sided MA representation involving past, current and future values of ε_t . In case $q > 0$, the processes considered no longer have a strictly stationary solution solely in terms of ε_t , but involve both \mathbf{X}_t and ε_t . That is,

$$y_t = \pi(L, L^{-1})[\varepsilon_t + \boldsymbol{\beta}' \mathbf{X}_t] = \sum_{j=-\infty}^{\infty} \pi_j z_{t-j}, \quad (4.7)$$

where $z_{t-j} = \varepsilon_{t-j} + \sum_{i=1}^q \beta_i x_{i,t-j}$ and $\pi(z, z^{-1})$ is an operator satisfying the relation $\pi(z, z^{-1})\phi(z)\varphi(z^{-1}) = 1$. Note that we replace the operator L by the complex variable z when considering the properties of polynomials. Similar to Gouriéroux and Jasiak (2016), we note that the polynomials $\phi(z)$ and $\varphi(z^{-1})$ are invertible and their inverses create infinite series in z and z^{-1} respectively, causing (4.7) to hold almost surely (see Brockwell and Davis, 1991, proposition 13.3.1 for more details).

We observe that y_t still has a two-sided MA-representation, but augmented with a second part involving linear combinations of past, current and future values of \mathbf{X}_t . Since $\beta_i x_{i,t}$ can be interpreted as a new series $\tilde{x}_{i,t}$ which is the series $x_{i,t}$ multiplied by a constant term β_i , y_t in (4.7) consists of two additive parts: (i) a two-sided MA representation and (ii) the sum of q processes $\tilde{x}_{i,t}$ that are passed through a two-sided linear filter with coefficients resulting from inverting the product $[\phi(z)\varphi(z^{-1})]$ to $\pi(z, z^{-1})$.⁶

Lemma 4.1. *From (4.3), we can construct the unobserved noncausal and causal components (u, v) similar to Lanne and Saikkonen (2011a) and Gouriéroux and Jasiak (2016) and obtain:*

$$u_t \equiv \phi(L)y_t \leftrightarrow \varphi(L^{-1})u_t - \boldsymbol{\beta}' \mathbf{X}_t = \varepsilon_t, \quad (4.8)$$

$$v_t \equiv \varphi(L^{-1})y_t \leftrightarrow \phi(L)v_t - \boldsymbol{\beta}' \mathbf{X}_t = \varepsilon_t. \quad (4.9)$$

In order to ensure identifiability of the parameter vector $\boldsymbol{\beta}$ and to prove consistency of the ML estimator, we make the following assumptions on \mathbf{X}_t :

⁶The effects of two-sided linear filters, with a focus on seasonal adjustment, on the identification of mixed causal-noncausal models is studied in Hecq et al. (2017c).

Assumptions. The processes in \mathbf{X}_t are assumed to be

(A1) ergodic, (strictly) stationary and strictly exogenous w.r.t. ε_t (\mathbf{X}_t and ε_t are independent stochastic processes);

(A2) mixed causal-noncausal with finite variance, i.e., $x_{i,t} = c_i + \sum_{j=-\infty}^{\infty} \rho_{i,j} \eta_{i,t-j}$ with $\eta_{i,t}$ strong white noise;

(A3) linearly independent.

Assumption (A1) is necessary for the Central Limit Theorem for m -dependent processes (Theorem 6.4.2 in Brockwell and Davis, 1991). Assumption (A2) defines the dynamic structure the processes in \mathbf{X}_t can assume, allowing for a two-sided moving process, purely causal and noncausal as well as mixed ARMA processes are allowed. By assumption (A3) these processes are linearly independent, which simplifies the proof for consistency of the ML estimator. This assumption can be relaxed, but we leave this for future research, as it is not restrictive for the empirical application considered in this chapter.

4.2.1 Simulation of MARX Processes

The filtered values defined in Lemma 4.1 establish a deterministic dynamic relationship between the unobserved components u_t and v_t , the exogenous variables \mathbf{X}_t and the process y_t , which can be used to simulate various MARX(r, s, q) series. Gouriéroux and Jasiak (2016) show extensively how to simulate MAR(r, s) processes and make use of the independence of specific blocks of u, v and y values. We extend their analysis to the MARX(r, s, q) case and show that the equivalence of different information sets still holds.

The main difficulty for generating MARX(r, s, q) with both $r, s \geq 1$ is the product of polynomials $\phi(z)\varphi(z^{-1})$. One cannot directly simulate such a process as simultaneously initial and terminal values are required. If the degree of (at least) one of the polynomials equals 0 (i.e., the purely causal, noncausal and static case), the problem is greatly simplified. We illustrate this by considering the MARX(0,1,1) model. In that case (4.3) reduces to:

$$y_t = \varphi_1 y_{t+1} + \beta_1 x_{1,t} + \varepsilon_t, \quad (4.10)$$

which can easily be simulated directly by generating a sequence of ε_t and choosing terminal values y_T and $x_{1,T}$.⁷ In the general MARX(r, s, q) setup, filtered values

⁷A burn-in period should be considered to delete dependence on terminal values.

are used to circumvent the problem. Defining $[\varphi(z^{-1})]^{-1} \equiv \delta(z^{-1})$, we can rewrite the second equality in (4.8) in the following way:

$$u_t = \sum_{j=0}^{\infty} \delta_j \left(\sum_{i=1}^q \beta_i x_{i,t+j} + \varepsilon_{t+j} \right) = \sum_{j=0}^{\infty} \delta_j z_{t+j}. \quad (4.11)$$

In a similar fashion, when we take $[\phi(z)]^{-1} \equiv \alpha(z)$, we obtain for v_t :

$$v_t = \sum_{j=0}^{\infty} \alpha_j \left(\sum_{i=1}^q \beta_i x_{i,t-j} + \varepsilon_{t-j} \right) = \sum_{j=0}^{\infty} \alpha_j z_{t-j}. \quad (4.12)$$

Using these expressions, $\text{MARX}(r, s, q)$ can be constructed directly by means of the definitions given in (4.8) and (4.9). That is, the causal and noncausal components (u, v) can be simulated independently and can be interpreted as a causal [noncausal] “error term” of a purely noncausal [causal] autoregression.

We characterize the simulation steps using (4.8). Note that the first equality $\phi(L)y_t = u_t$ appears like a conventional causal autoregressive process. In order to simulate such a process, one needs r starting values, say $y_{-1}^*, \dots, y_{-r}^*$, to create the first value y_1 . Additionally, one needs the value v_1 which is usually a draw from a desired distribution in the case of a causal autoregressive process. In the MARX case, however, v_1 is represented as a linear combination of current and future values of ε_t and \mathbf{X}_t as can be seen in (4.11). If we consider a truncation m sufficiently large for the infinite sum, v_1 can be constructed by simulating long paths of ε_t and \mathbf{X}_t such that z_1 up to z_{1+m} are available. Now that y_1 is generated, it can be used to construct the next value y_2 . In general, the following two steps can be used to simulate an MARX processes based on (4.8):

1. Generate a path of length $(T + m)$ for ε_t and \mathbf{X}_t and simulate the values of u_t using a truncated version of (4.11).
2. Create *starting* values y_{-1}^* up to y_{-r}^* and simulate the process y_t like a conventional causal autoregressive process.

It is also possible to base the simulation on (4.9), then one typically generates the series “backwards”. In that case, the following steps are needed:

1. Generate a path of length $(T + m)$ for ε_t and \mathbf{X}_t and simulate the values of v_t using a truncated version of (4.12).

2. Create *terminal* values y_{T+1}^* up to y_{T+s}^* and simulate the process y_t like a noncausal autoregressive process (see e.g., Gouriéroux and Jasiak, 2016).

Example 4.1. An MARX(1,1,1) can be simulated according to

$$y_t = \varphi_1 y_{t+1} + \sum_{j=0}^{\infty} \phi^j (\beta_1 x_{1,t-j} + \varepsilon_{t-j}), \quad (4.13)$$

where a truncation m sufficiently large has to be considered for the infinite sum.⁸ The simulation of this process consists of two steps. Firstly, simulate a long path of length $T + m$ for $x_{1,t}$ and ε_t . Accordingly, one can construct the sequence v_t for $t = 1, \dots, T$. Secondly, choose a terminal value, say y_{T+1}^* . Using (4.13), we are going to simulate the series “backwards”. That is, to simulate y_T , we put the terminal value y_{T+1}^* for y_{t+1} in (4.13) and use the value v_T simulated in the first step. Now that y_T becomes available, it can be used in combination with v_{T-1} to construct y_{T-1} . We continue this procedure until we reach y_1 .

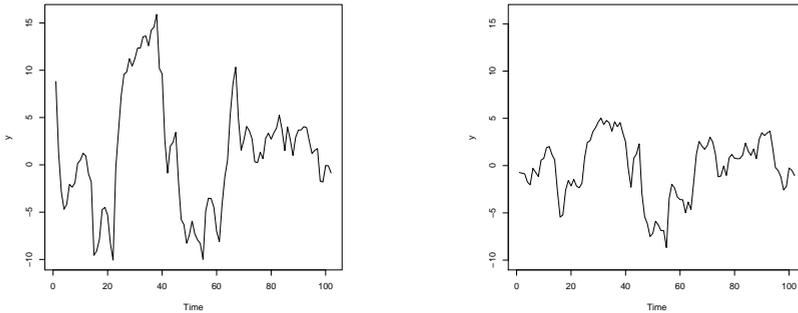


Figure 4.1: A simulated MARX process (left) and the same process without exogenous variable (right)

Figure 4.1 shows simulated paths of an MARX(1,1,1) and MAR(1,1) process with $[\phi_1, \varphi_1]' = [0.3, 0.9]'$, $\beta_1 = 0.3$, $x_{1,t} \stackrel{iid}{\sim} t(1, 0)$ and $\varepsilon_t \stackrel{iid}{\sim} t(3, 0)$ with $t(\nu, \sigma)$ being the Student’s t distribution with degrees of freedom parameter ν and scale parameter σ . The truncation m is set at 10,000. It can be seen that both processes generally move similarly with the major exception that the MARX process contains

⁸Since $\deg(\phi(z)) = 1$ in this instance, it is straightforward to compute the inverse of this polynomial. For more complicated polynomials, one could compute a companion matrix to find its inverse.

more peaks and troughs, which are also more extreme in comparison. This is due to the choice of $x_{1,t}$, which is chosen to be standard Cauchy distributed for expository purposes. Hence, the MARX specification takes into account shocks that cannot be explained by past, current and future values of the dependent variable, but which are present because of major changes in explanatory exogenous variables at some specific points in time. In order to justify the simulation method as outlined above, we present the following proposition that shows the equivalence of various information sets. A proof of this proposition is in the Appendix 4.C.

Proposition 4.1. For an MARX(r, s, q) model, the following information sets are equivalent:

- (i) $(y_1, \dots, y_T, \mathbf{X}_{r+1}, \dots, \mathbf{X}_{T-s})$
- (ii) $(y_1, \dots, y_r, u_{r+1}, \dots, u_T, \mathbf{X}_{r+1}, \dots, \mathbf{X}_{T-s})$
- (iii) $(v_1, \dots, v_{T-s}, y_{T-s+1}, \dots, y_T, \mathbf{X}_{r+1}, \dots, \mathbf{X}_{T-s})$
- (iv) $(y_1, \dots, y_r, \varepsilon_{r+1}, \dots, \varepsilon_{T-s}, u_{T-s+1}, \dots, u_T)$
- (v) $(v_1, \dots, v_r, \varepsilon_{r+1}, \dots, \varepsilon_{T-s}, y_{T-s+1}, \dots, y_T)$
- (vi) $(v_1, \dots, v_r, \varepsilon_{r+1}, \dots, \varepsilon_{T-s}, u_{T-s+1}, \dots, u_T)$

Additionally, the following information sets are equivalent:

- (i') (y_1, \dots, y_T)
- (ii') $(y_1, \dots, y_r, u_{r+1}, \dots, u_T)$
- (iii') $(v_1, \dots, v_{T-s}, y_{T-s+1}, \dots, y_T)$

From this we deduce that the three sets of variables (v_1, \dots, v_r) , $(\varepsilon_{r+1}, \dots, \varepsilon_{T-s})$ and (u_{T-s+1}, \dots, u_T) are also independent in the MARX setup. Similar to Gouriéroux and Jasiak (2016), we can still interpret $(v_1, \dots, v_r) [(u_{T-s+1}, \dots, u_T)]$ as the initial [terminal] conditions that determine the path of process y_t over the period $1, \dots, T$.

4.2.2 Identifiability

Identifiability of mixed causal-noncausal models has received a lot of attention in the literature. Since the Gaussian distribution is fully characterized by its

autocovariance function and spectral density (which are symmetric), it is well-known that forward- and backward-looking behavior cannot be distinguished (see e.g., Breidt et al., 1991). For this reason, estimation methods based on solely second order properties of the data (like e.g., OLS and Gaussian MLE) cannot be used to identify such models. As the inclusion of exogenous variables introduces the presence of cross-covariances, an interesting property of MARX models is the possibility to identify models even under Gaussianity.

For simplicity, let us focus on the case of one exogenous regressor x_t . In order to illustrate the identifiability, we consider the cross-covariances $\mathbb{E}(y_t x_{t-k})$ and $\mathbb{E}(y_t x_{t+k})$ for $k \geq 1$ in three cases:

- (1) y_t is a mixed causal-noncausal process as in (4.3);
- (2) y_t is a purely causal process as in (4.4);
- (3) y_t is a purely noncausal process as in (4.5).

We will argue that, in most cases, identification is possible, as the cross-covariances are generally different in the three cases considered. The properties of x_t play a vital role in this. Using the assumptions (A1)-(A3), the notation $\gamma_x(k)$ to denote the k th order autocovariance of x_t , $[\phi(z)\varphi(z^{-1})]^{-1} \equiv \pi(z, z^{-1})$, $[\phi(z)]^{-1} \equiv \alpha(z)$ and $[\varphi(z^{-1})]^{-1} = \delta(z^{-1})$, we can find closed form expressions for the two cross-covariances in all three cases. These can be found in Table 4.1.

	Model		
	Mixed causal-noncausal	Causal	Noncausal
$\mathbb{E}(y_t x_{t-k})$	$\beta \sum_{j=-\infty}^{\infty} \pi_j \gamma_x(j-k)$	$\beta \sum_{j=0}^{\infty} \alpha_j \gamma_x(j-k)$	$\beta \sum_{j=0}^{\infty} \delta_j \gamma_x(j+k)$
$\mathbb{E}(y_t x_{t+k})$	$\beta \sum_{j=-\infty}^{\infty} \pi_j \gamma_x(j+k)$	$\beta \sum_{j=0}^{\infty} \alpha_j \gamma_x(j+k)$	$\beta \sum_{j=0}^{\infty} \delta_j \gamma_x(j-k)$

Table 4.1: Cross-covariances for different model specifications

The expressions in Table 4.1 appear very similar, but an important observation can be made. Note that, in the case of the mixed causal-noncausal model, the variance of x_t is included in both $\mathbb{E}(y_t x_{t-k})$ (by choosing $j = k$) and $\mathbb{E}(y_t x_{t+k})$ (by choosing $j = -k$). In contrast, we see for the causal model that only $\mathbb{E}(y_t x_{t-k})$ includes the variance of x_t while $\mathbb{E}(y_t x_{t+k})$ consists only of autocovariances of x_t at order k and higher. For the noncausal model, the results are exactly opposite: $\mathbb{E}(y_t x_{t+k})$ consists of the variance of x_t while $\mathbb{E}(y_t x_{t-k})$ is a linear combination

of autocovariances of x_t excluding the variance. Hence, if we assume x_t to be *iid* distributed, identification of the model can easily be achieved by computing the cross-covariances in the following way:

- If both $\mathbb{E}(y_t x_{t-k})$ and $\mathbb{E}(y_t x_{t+k})$ are nonzero, the process is mixed causal-noncausal.
- If $\mathbb{E}(y_t x_{t-k})$ is nonzero, while $\mathbb{E}(y_t x_{t+k})$ is zero, the process is purely causal.
- If $\mathbb{E}(y_t x_{t+k})$ is nonzero, while $\mathbb{E}(y_t x_{t-k})$ is zero, the process is purely noncausal.

For series x_t with a dynamic structure, we cannot obtain the same result. However, a similar conclusion can be reached by taking a closer look at the cross-covariances in Table 4.1. For illustration purposes, let us focus on models with autoregressive order equal to 1. This means that the two candidate models for y_t are the MARX(1,0,1) and MARX(0,1,1). Assume the latter one is the true data generating process, which means $\delta_j = \varphi_1^j$. An identification problem may arise when $\mathbb{E}(y_t x_{t-k})$ and $\mathbb{E}(y_t x_{t+k})$ take on the same value, which is exactly the case when

$$\mathbb{E}(y_t x_{t+k}) - \mathbb{E}(y_t x_{t-k}) = \beta \left[\sum_{i=-k}^{k-1} \varphi_1^{i+k} \gamma_x(i) + \sum_{i=k}^{\infty} (\varphi_1^{i+k} - \varphi_1^{i-k}) \gamma_x(i) \right] \quad (4.14)$$

reduces to zero and $\alpha_j = \delta_j (= \varphi_1^j$ in this case). By symmetry, we have that also for the causal model the difference between the two cross-covariances equals zero. In fact, the four cross-covariances for the causal and noncausal model are all exactly the same in this situation. However, in theory, we have $\alpha_j \neq \delta_j$, which ensures that the process can still be distinguished from a purely causal process. Let us consider an example which characterizes this identification through the coefficients.

Example 4.2. Let y_t follow an MARX(0,1,1) where x_t is a process with $\gamma_x(0) = -\varphi_1 \gamma_x(1)$, $\gamma_x(1) \neq 0$, while $\gamma_x(k) = 0$ for all $k > 1$. Then it can be observed that (4.14) reduces to zero. In the MARX(0,1,1) case: $\mathbb{E}(y_t x_{t-k}) = \beta \gamma_x(1) = \mathbb{E}(y_t x_{t+k})$, using $\varphi_1^0 = 1$. In the case of a causal MARX(1,0,1): $\mathbb{E}(y_t x_{t-k}) = \beta[(1 + \varphi_1^{-2})\gamma_x(1) + \varphi_1^{-1}\gamma_x(0)] = \beta\varphi_1^{-2}\gamma_x(1)$, while $\mathbb{E}(y_t x_{t+k}) = \beta\gamma_x(1)$.

Note that since $\gamma_x(0) = -\varphi_1 \gamma_x(1)$, the difference in coefficients ($\alpha_j = \varphi_1^{-j}$ while $\delta_j = \varphi_1^j$) is the key element that makes one cross-covariance (i.e. $\mathbb{E}(y_t x_{t-k})$ in the causal model) different, as not the same terms cancel out due to a difference in

coefficients. In contrast, if it had been the case that $\alpha_j = \delta_j = \varphi_1^j$ in this example, all cross-covariance would have reduced to the same value and no identification could be achieved.

The example considered is a rather exceptional situation. Hence, in theory, the difference in coefficients between mixed causal-noncausal models and the purely causal and noncausal models should establish differences in the cross-covariances that could be exploited to track down the right model. The possibility to identify models in practice is less straightforward when x_t exhibits dynamics, as estimation uncertainty in both coefficients and the autocovariances makes it difficult to compare the estimated cross-covariances with the theoretical ones, which are often not even known.

The discussion for the MARX(1,0,1) and MARX(0,1,1) above can be generalized to processes with higher order dynamics and more than one exogenous regressor. Not only the dynamics of the exogenous regressors are of importance, but also their marginal distributions. In case they possess an infinite unconditional variance, the cross-covariances might not be defined.⁹ In this chapter, assumption (A2) excludes those processes from the analysis and thus we do not consider them here. Another consideration, which is outside the scope of this chapter, is the situation in which the exogenous regressors are linearly correlated.

4.3 Maximum Likelihood Estimation

Maximum likelihood estimation of noncausal autoregressive models has been studied by Breidt et al. (1991), Andrews et al. (2006) and Lanne and Saikkonen (2011a).¹⁰ They show that ML estimators are consistent and asymptotically normal under general conditions. This section builds on these results and establishes similar results for mixed causal-noncausal autoregressive models with exogenous regressors.

⁹This need not impose a problem, since Gouriéroux and Zakoïan (2016) show that sample autocorrelations are consistent for noncausal processes with α -stable error term, $\alpha \in (0, 2)$, despite the fact that the theoretical autocorrelations do not exist.

¹⁰Breidt et al. (1991) specify a noncausal model as a conventional autoregressive model that has roots inside the unit circle, while Andrews et al. (2006) consider all-pass models which are widely used in fitting noncausal autoregressions. Lanne and Saikkonen (2011a) have a similar model setup to (4.3), the only difference being the exclusion of exogenous variables \mathbf{X}_t .

Similar to Lanne and Saikkonen (2011a), we assume that the density function $f(x; \boldsymbol{\lambda})$ satisfies the regularity conditions of Andrews et al. (2006).¹¹ The permissible parameter space of $\boldsymbol{\lambda}$, denoted by $\boldsymbol{\Lambda}$, is some subset of \mathbb{R}^d . The scale parameter only takes positive values, i.e., $\sigma > 0$. The permissible space of the parameters $\boldsymbol{\phi}$ and $\boldsymbol{\varphi}$ is defined by the stationarity condition (4.6). Using the independence of the blocks (v_1, \dots, v_r) , $(\varepsilon_{r+1}, \dots, \varepsilon_{T-s})$ and (u_{T-s+1}, \dots, u_T) , it is shown in Appendix 4.B that the density of the process y_t can be written as the product of the densities of these three variables. However, since (the densities of) the first and third block do not depend on sample size T , we can approximate the density of y_t by the density of the second block. Replacing ε_t by the left-hand side of (4.3) and taking logs, we obtain the following log-likelihood function:

$$\begin{aligned} L_T(\boldsymbol{\theta}) &= \sum_{t=r+1}^{T-s} \ln f_{\sigma}(\boldsymbol{\phi}(L)\boldsymbol{\varphi}(L^{-1})y_t - \boldsymbol{\beta}'\mathbf{X}_t; \boldsymbol{\lambda}) \\ &= \sum_{t=r+1}^{T-s} g_t(\boldsymbol{\theta}), \end{aligned} \tag{4.15}$$

where $\boldsymbol{\theta} = [\boldsymbol{\phi}', \boldsymbol{\varphi}', \boldsymbol{\beta}', \boldsymbol{\lambda}', \sigma]'$. For convenience, we denote the ‘approximate’ sample size used to compute the log-likelihood $(T - p)$ by n . We can use the definition of the filtered values as in (4.8) and (4.9) to write the series u_t and v_t as functions of the parameters, i.e., $u_t(\boldsymbol{\phi})$ and $v_t(\boldsymbol{\varphi})$. Then we can characterize $g_t(\boldsymbol{\theta})$ as follows:

$$\begin{aligned} g_t(\boldsymbol{\theta}) &= \ln f(\sigma^{-1}(v_t(\boldsymbol{\varphi}) - \phi_1 v_{t-1}(\boldsymbol{\varphi}) - \dots - \phi_r v_{t-r}(\boldsymbol{\varphi}) - \boldsymbol{\beta}'\mathbf{X}_t); \boldsymbol{\lambda}) - \ln(\sigma) \\ &= \ln f(\sigma^{-1}(u_t(\boldsymbol{\phi}) - \varphi_1 u_{t+1}(\boldsymbol{\phi}) - \dots - \varphi_s u_{t+s}(\boldsymbol{\phi}) - \boldsymbol{\beta}'\mathbf{X}_t); \boldsymbol{\lambda}) - \ln(\sigma), \end{aligned}$$

where we also used $f_{\sigma}(x; \boldsymbol{\lambda}) = \sigma^{-1} f(\sigma^{-1}x; \boldsymbol{\lambda})$ (definition of density). Maximizing $L_T(\boldsymbol{\theta})$ over permissible values of $\boldsymbol{\theta}$ gives an approximate ML estimator of $\boldsymbol{\theta}$. We assume for now that the orders r and s are known. Denote the true value of $\boldsymbol{\theta}$ by $\boldsymbol{\theta}_0$ (and similarly for its components). Furthermore, assume that $\boldsymbol{\lambda}_0$, the true value of $\boldsymbol{\lambda}$, is an interior point of $\boldsymbol{\Lambda}$.

¹¹The regularity conditions of Andrews et al. (2006) will henceforth be assumed. Densities that satisfy these conditions include a rescaled t -density and a weighted average of Gaussian densities.

4.3.1 Asymptotic Properties of the AML Estimator

We first consider the score of θ evaluated at true parameter values. Define $\mathbf{V}_{t-1} = [v_{t-1}, \dots, v_{t-r}]'$ and $\mathbf{U}_{t+1} = [u_{t+1}, \dots, u_{t+s}]'$, where u_t and v_t are defined in terms of true parameter values, i.e., $u_t = \sum_{j=0}^{\infty} \delta_{0j} (\sum_{i=1}^q \beta_{0i} x_{i,t+j} + \varepsilon_{t+j})$ and $v_t = \sum_{j=0}^{\infty} \alpha_{0j} (\sum_{i=1}^q \beta_{0i} x_{i,t-j} + \varepsilon_{t-j})$. By direct differentiation of (4.15), we obtain:

$$\frac{\partial}{\partial \phi} g_t(\theta_0) = -\frac{f'(\sigma_0^{-1} \varepsilon_t; \lambda_0)}{\sigma_0 f(\sigma_0^{-1} \varepsilon_t; \lambda_0)} \mathbf{V}_{t-1}, \quad \frac{\partial}{\partial \varphi} g_t(\theta_0) = -\frac{f'(\sigma_0^{-1} \varepsilon_t; \lambda_0)}{\sigma_0 f(\sigma_0^{-1} \varepsilon_t; \lambda_0)} \mathbf{U}_{t+1},$$

and

$$\frac{\partial}{\partial \beta} g_t(\theta_0) = -\frac{f'(\sigma_0^{-1} \varepsilon_t; \lambda_0)}{\sigma_0 f(\sigma_0^{-1} \varepsilon_t; \lambda_0)} \mathbf{X}_t,$$

where $f'(x; \lambda) = \partial f(x; \lambda) / \partial x$ and use has been made of the fact that $\varphi_0(L^{-1})u_t - \beta_0' \mathbf{X}_t = \varepsilon_t = \phi_0(L)v_t - \beta_0' \mathbf{X}_t$. Similarly, for the distributional parameters:

$$\frac{\partial}{\partial \sigma} g_t(\theta_0) = -\sigma_0^2 \left(\frac{f'(\sigma_0^{-1} \varepsilon_t; \lambda_0)}{f(\sigma_0^{-1} \varepsilon_t; \lambda_0)} + \sigma_0 \right), \quad \frac{\partial}{\partial \lambda} g_t(\theta_0) = \frac{1}{f(\sigma_0^{-1} \varepsilon_t; \lambda_0)} \frac{\partial}{\partial \lambda} f(\sigma_0^{-1} \varepsilon_t; \lambda_0).$$

The following lemma presents the asymptotic distribution of the score vector. Define

$$\mathcal{J} = \int \frac{(f'(x; \lambda_0))^2}{f(x; \lambda_0)} dx > 1 \quad \text{and} \quad \mathcal{I} = \int x^2 \frac{(f'(x; \lambda_0))^2}{f(x; \lambda_0)} dx - 1,$$

where the first inequality follows from Remark 2 in Andrews et al. (2006). Furthermore set

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{bmatrix}.$$

The matrix Σ is symmetric and has the matrices $\Sigma_{11} = \sigma_0^{-2} \mathcal{J} \Gamma_V$, $\Sigma_{22} = \sigma_0^{-2} \mathcal{J} \Gamma_U$ and $\Sigma_{33} = \sigma_0^{-2} \mathcal{J} \Gamma_X$ on the diagonal, where Γ_V and Γ_U are the autocovariance matrices of \mathbf{V}_{t-1} and \mathbf{U}_{t+1} respectively. Γ_X is the cross-covariance matrix of the q processes in \mathbf{X}_t which is diagonal under the assumption of linear independence

between processes in \mathbf{X}_t . Σ_{12} is a $(r \times s)$ matrix where the (i, j) th element equals:

$$\sum_{t=0}^{\infty} \alpha_t \delta_{t+i-j} + \tilde{\mathcal{J}} \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} \alpha_a \delta_b \sum_{m=1}^q \beta_m^2 \gamma_{x_m}(i+j+a+b)$$

The Σ_{13} matrix has size $(r \times q)$ with the (i, j) th element equal to

$$\beta_j \sigma^{-2} \mathcal{J} \sum_{a=0}^{\infty} \alpha_a \gamma_{x_j}(i+a),$$

while for Σ_{23} this element is

$$\beta_j \sigma^{-2} \mathcal{J} \sum_{b=0}^{\infty} \delta_b \gamma_{x_j}(i+b),$$

Note that the summands involve only i and not j , as only the former denotes the lag or lead considered for v_t and u_t respectively. In contrast, j ($= 1, \dots, q$) runs over all exogenous variables $x_{1,t}, \dots, x_{q,t}$. Finally define the $(d+1) \times (d+1)$ matrix

$$\Omega = \begin{bmatrix} \omega_{\sigma}^2 & \omega_{\sigma\lambda} \\ \omega_{\lambda\sigma} & \Omega_{\lambda\lambda} \end{bmatrix},$$

where

$$\Omega_{\lambda\lambda} = \int \frac{1}{f(x; \lambda_0)} \left(\frac{\partial}{\partial \lambda} f(x; \lambda) \right) \left(\frac{\partial}{\partial \lambda} f(x; \lambda) \right)' dx,$$

$$\omega_{\lambda\sigma} = -\sigma_0 \int x \frac{f'(x; \lambda_0)}{f(x; \lambda_0)} \frac{\partial}{\partial \lambda} f(x; \lambda_0) dx = \omega'_{\sigma\lambda},$$

and

$$\omega_{\sigma}^2 = \omega_0^{-2} \mathcal{I}.$$

Lemma 4.2. *If conditions (A1)-(A7) of Andrews et al. (2006) hold, then*

$$\frac{1}{\sqrt{n}} \sum_{t=r+1}^{T-s} \frac{\partial}{\partial \theta} g_t(\theta_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \text{diag}(\Sigma, \Omega)).$$

Moreover, the matrices Σ and Ω are positive definite.

The block diagonality of the covariance matrix of the limiting distribution follows directly from the formulation of the model in terms of both lag and lead operator. Hence, this result follows directly from Lanne and Saikkonen (2011a) and Andrews et al. (2006). This ensures that both the autoregressive and exogenous (variables) parameters are orthogonal to the distributional parameters. The positive definiteness of $\mathbf{\Omega}$ is assumed through condition (A6) of Andrews et al. (2006). The positive definiteness of $\mathbf{\Sigma}$ follows, similar to the MAR case, from the condition $\mathcal{J} > 1$.

Theorem 4.1. *If conditions (A1)-(A7) of Andrews et al. (2006) hold, there exists a sequence of local maximizers $\hat{\boldsymbol{\theta}} = [\hat{\boldsymbol{\phi}}', \hat{\boldsymbol{\varphi}}', \hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\lambda}}', \hat{\sigma}]'$ of $L_T(\boldsymbol{\theta})$ in (4.15) such that*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \text{diag}(\mathbf{\Sigma}^{-1}, \mathbf{\Omega}^{-1})).$$

4.3.2 Computing the Covariance Matrix

Block diagonality of the covariance matrix in Theorem 4.1 shows that the approximate ML estimators of the model parameters $[\boldsymbol{\phi}', \boldsymbol{\varphi}', \boldsymbol{\beta}']'$ and the distributional parameters $[\sigma, \boldsymbol{\lambda}']'$ are asymptotically independent. The computation of this covariance matrix is of interest when one wants to compute (approximate) standard errors of the parameters for inference (e.g., confidence levels, hypothesis testing). A conventional estimator is based on the Hessian of the log-likelihood but nonlinear optimization of this function often involves complicated gradient based numerical methods. As these procedures are relatively unstable in certain settings, we provide an alternative way to compute the standard errors of the parameters of both the autoregressive and exogenous variables for Student's t -MLE and the LAD estimator.

Student's t Maximum Likelihood Estimation

Similar to Hecq et al. (2016), we can characterize the asymptotic distribution of the Student's t -MLE and LAD estimated parameters in the finite variance framework. If $\nu > 2$, the MLE is \sqrt{n} -consistent and asymptotically normal. Define the $(n \times 1)$ series $\mathbf{u} \equiv \mathbf{U}_t^* = [u_{r+1}, \dots, u_{T-s}]'$ up to $\mathbf{U}_{t+s}^* = [u_{r+s+1}, \dots, u_T]'$, $\mathbf{V}_{t-r}^* = [v_1, \dots, v_{T-p}]'$ up to $\mathbf{v} \equiv \mathbf{V}_t^* = [v_{r+1}, \dots, v_{T-s}]'$, $\mathbf{X}_{i,t} = [x_{i,r+1}, \dots, x_{i,T-s}]'$ and $\boldsymbol{\varepsilon} = [\varepsilon_{r+1}, \dots, \varepsilon_{T-s}]'$. We construct the matrices $\mathbf{Z} = [\mathbf{U}_{t+1}^*, \dots, \mathbf{U}_{t+s}^*, \mathbf{X}_{1,t}, \dots, \mathbf{X}_{q,t}]$ and similarly $\mathbf{Q} = [\mathbf{V}_{t-1}^*, \dots, \mathbf{V}_{t-r}^*, \mathbf{X}_{1,t}, \dots, \mathbf{X}_{q,t}]$, which are of dimensions $(n \times (s+q))$ and $(n \times (r+q))$ respectively. Using this notation, we can write the autoregressions

defined in (4.8) and (4.9) in matrix notation as follows:

$$\mathbf{u} = \mathbf{Z}\boldsymbol{\zeta} + \boldsymbol{\varepsilon}, \quad (4.16)$$

$$\mathbf{v} = \mathbf{Q}\boldsymbol{\xi} + \boldsymbol{\varepsilon}, \quad (4.17)$$

with $\boldsymbol{\zeta} = [\boldsymbol{\varphi}', \boldsymbol{\beta}']' \in \mathbb{R}^{s+q}$ and $\boldsymbol{\xi} = [\boldsymbol{\phi}', \boldsymbol{\beta}']' \in \mathbb{R}^{r+q}$.

Then, conditional on the unobserved causal and noncausal components discussed above, it can be shown that in the case of an MARX(r, s, q) model

$$\sqrt{n}(\hat{\boldsymbol{\zeta}}_{ML} - \boldsymbol{\zeta}_0) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \frac{\nu+3}{\nu+1}\sigma^2\boldsymbol{\Upsilon}_\phi^{-1}\right), \quad (4.18)$$

$$\sqrt{n}(\hat{\boldsymbol{\xi}}_{ML} - \boldsymbol{\xi}_0) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \frac{\nu+3}{\nu+1}\sigma^2\boldsymbol{\Upsilon}_\varphi^{-1}\right), \quad (4.19)$$

holds. We use the notation $\boldsymbol{\Upsilon}_\varphi = \mathbb{E}[\mathbf{Q}'\mathbf{Q}]$ and $\boldsymbol{\Upsilon}_\phi = \mathbb{E}[\mathbf{Z}'\mathbf{Z}]$, where $\boldsymbol{\varphi}$ and $\boldsymbol{\phi}$ signify the relation between the unobserved values \mathbf{u}, \mathbf{v} and \mathbf{y} as defined in (4.8)-(4.9). These quantities can be estimated consistently by $(1/n)\sum_{i=1}^n \mathbf{Q}'_i \mathbf{Q}_i$ and $(1/n)\sum_{i=1}^n \mathbf{Z}'_i \mathbf{Z}_i$, where \mathbf{Q}_i [resp. \mathbf{Z}_i] denotes the i th row of the matrix \mathbf{Q} [resp. \mathbf{Z}]. For large ν , i.e., $\nu \rightarrow \infty$, l_y approaches the Gaussian (log)-likelihood, and the model parameters cannot be consistently estimated anymore.

Least Absolute Deviation Estimation

Since the LAD can be used as an initial estimator for $\boldsymbol{\phi}$ and $\boldsymbol{\varphi}$ (Lanne and Saikkonen, 2011a) and is found to outperform Student's t -MLE in certain instances (Hecq et al., 2016), we also present the asymptotic distribution of the model parameters for this estimation method. If ε_t is a sequence of *iid* random variables with mean zero, median zero, finite variance and probability density function $f_\varepsilon(\varepsilon_t; \boldsymbol{\lambda})$ that is continuous in a neighborhood of zero, the LAD estimator is \sqrt{n} -consistent and asymptotically normal (Wu and Davis, 2010). Following Hecq et al. (2016) and using the model specifications in (4.16)-(4.17), it follows that, conditional on the unobserved causal and noncausal components,

$$\sqrt{n}(\boldsymbol{\zeta}_{LAD} - \boldsymbol{\zeta}_0) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \frac{1}{4f_\varepsilon^2(0)}\boldsymbol{\Upsilon}_\phi^{-1}\right), \quad (4.20)$$

$$\sqrt{n}(\boldsymbol{\xi}_{LAD} - \boldsymbol{\xi}_0) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \frac{1}{4f_\varepsilon^2(0)}\boldsymbol{\Upsilon}_\varphi^{-1}\right), \quad (4.21)$$

where $f_\varepsilon(0)$ can be estimated by a logistic kernel. The LAD estimator can be interpreted as a maximum likelihood estimator for which the error term follows a Laplacian distribution. It should be noted that the density of this distribution does not satisfy the regularity conditions of Andrews et al. (2006).

4.4 Simulation Study

By means of Monte Carlo simulations, we investigate three different cases of interest: (i) the performance of the MLE for MARX processes, (ii) the identifiability of MARX models under Gaussianity and (iii) a model selection procedure for MARX models. Each table in our simulation study reports results for 1000 replications.

4.4.1 Performance MLE for MARX

To assess the performance of the maximum likelihood estimator, we take the following MARX(1,1,1) as data generating process (DGP):

$$(1 - \phi_1 L)(1 - \varphi_1 L^{-1})y_t - \beta_1 x_{1,t} = \varepsilon_t, \quad (4.22)$$

where $\phi_1 = 0.3$, $\varphi_1 = 0.5$ and $\beta_1 = 0.3$. The error term ε_t follows a t -distribution with 3 degrees of freedom.¹² $x_{1,t}$ will follow different specifications:

- (1) $x_{1,t} \stackrel{iid}{\sim} t(5, 0)$,
- (2) $x_{1,t} \stackrel{iid}{\sim} \mathcal{N}(0, 1)$,
- (3) $x_{1,t} \stackrel{iid}{\sim} C(0, 1)$,
- (4) $x_{1,t}$ follows an AR(1) process: $x_{1,t} = 0.6x_{1,t-1} + \varepsilon_t$ where $\varepsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, 5)$.

Table 4.2 reports the mean and standard deviations of the estimated parameters by MLE over all simulations. It can be seen that different specifications for $x_{1,t}$ only introduce relatively small differences. Most noticeably, the standard deviations of the parameters are larger for the first two cases especially when T is small. This can be due to the fact that both the $t(5, 0)$ and $\mathcal{N}(0, 1)$ distribution do not generate large outliers in $x_{1,t}$, which makes it more difficult to disentangle

¹²The same simulation study was performed for infinite variance cases, e.g., $\varepsilon_t \stackrel{iid}{\sim} t(2, 0)$. Similar to results in Hecq et al. (2016) for the MAR model, the simulation study suggests the fatter the tails of the error distribution, the more accurate estimation for both the coefficients and the distributional parameters of the MARX model.

their contribution to the series from that of lags and leads of y_t . The means of the estimated parameters also lie further away from the true value when compared to the other specifications, but are still very close. For all four specifications, the most difficult parameter to estimate is ν , which has a very large standard deviation for $T = 50$. For larger T , the standard deviations decrease rapidly. In all cases, the estimated mean over all parameters becomes more accurate and standard deviations decline as T grows large, which suggests the consistency of the maximum likelihood estimator in this framework.

		Specification for $x_{1,t}$							
		$x_t \stackrel{iid}{\sim} t(5, 0)$		$x_t \stackrel{iid}{\sim} \mathcal{N}(0, 1)$		$x_t \stackrel{iid}{\sim} C(0, 1)$		$x_t \sim \text{AR}(1)$	
T	Parameter	Mean	Std. dev	Mean	Std. dev	Mean	Std. dev	Mean	Std. dev
50	ϕ_1	0.313	0.168	0.309	0.168	0.297	0.073	0.291	0.078
	φ_1	0.461	0.164	0.469	0.156	0.491	0.072	0.495	0.065
	β_1	0.305	0.158	0.306	0.187	0.301	0.037	0.304	0.037
	ν	5.022	7.790	5.188	8.935	5.158	7.936	5.750	11.368
100	ϕ_1	0.302	0.110	0.301	0.113	0.300	0.039	0.296	0.052
	φ_1	0.489	0.103	0.484	0.106	0.497	0.036	0.499	0.042
	β_1	0.301	0.104	0.306	0.135	0.300	0.018	0.300	0.024
	ν	3.477	1.460	3.388	1.463	3.494	1.594	3.659	3.072
500	ϕ_1	0.301	0.037	0.300	0.040	0.300	0.016	0.299	0.022
	φ_1	0.497	0.034	0.498	0.036	0.500	0.008	0.500	0.017
	β_1	0.300	0.043	0.302	0.056	0.300	0.004	0.300	0.010
	ν	3.053	0.352	3.070	0.386	3.069	0.413	3.056	0.382
1000	ϕ_1	0.300	0.026	0.300	0.025	0.300	0.005	0.299	0.015
	φ_1	0.499	0.023	0.500	0.024	0.500	0.004	0.500	0.013
	β_1	0.300	0.030	0.300	0.038	0.300	0.002	0.300	0.007
	ν	3.039	0.254	3.033	0.244	3.027	0.281	3.031	0.258

Table 4.2: Finite sample properties of the ML estimator for an MARX(1,1,1) with $\varepsilon_t \stackrel{iid}{\sim} t(3, 0)$

4.4.2 Are MARX Models Identifiable Under Gaussianity?

In Section 4.2.2, we discussed the identifiability of MARX models even when the error term is normally distributed. To evaluate this important theoretical feature, we consider the purely noncausal data generating process as defined in (4.10) with both ε_t and $x_{1,t} \stackrel{iid}{\sim} \mathcal{N}(0, 1)$. We examine various combinations of parameter values for φ_1 and β_1 . To the simulated series y_t we fit both a correctly specified model, i.e., the MARX(0,1,1) and a misspecified autoregressive model with exogenous regressors, i.e. the MARX(1,0,1). The estimation method used is Gaussian MLE.¹³

¹³Due to condition (A5) of Andrews et al. (2006), consistency of Gaussian MLE for the MARX was not shown in Section 4.4.3 of this chapter. However, the consistency of Gaussian

We select the model that has the highest value for the log-likelihood function at the estimated parameters. Table 4.3 shows the percentages with which the correct model is chosen for different parameter values for φ_1 and β_1 .

We observe that the correct model is selected in approximately 50% of the cases when $\beta_1 = 0$ (irrespective of the value for φ_1). This is exactly in line with our expectations, as purely causal, mixed and purely noncausal autoregressive models cannot be identified by Gaussian MLE. Another special case arises when $\varphi_1 = 0$, as there is no autoregressive part present in the model. When $\beta_1 = 0$, we have a strong white noise and thus the correct model specification is not among the options. This is also the case for $\beta_1 \neq 0$, which causes the DGP to become a static regression. In both instances, both the MARX(1,0,1) and MARX(0,1,1) are chosen with roughly equal frequencies.

φ_1	β_1	$T = 50$	$T = 100$	$T = 200$	$T = 500$
0	0	49.0	51.9	49.1	50.0
	0.3	49.6	51.1	49.8	49.4
	0.7	49.4	50.5	49.3	49.1
0.1	0	50.1	51.9	49.2	50.0
	0.3	50.9	54.6	57.8	63.0
	0.7	55.5	61.3	68.6	82.7
0.4	0	52.2	52.4	50.3	50.8
	0.3	67.3	77.5	85.7	96.7
	0.7	88.7	96.6	99.6	100.0
0.6	0	53.4	53.0	50.4	51.3
	0.3	74.3	86.9	94.9	99.8
	0.7	96.0	99.7	100.0	100.0
0.8	0	50.9	50.2	48.6	51.4
	0.3	79.3	92.0	97.9	100.0
	0.7	98.9	100.0	100.0	100.0
0.9	0	50.4	48.9	47.5	51.9
	0.3	78.4	92.7	98.8	100.0
	0.7	99.1	99.9	100.0	100.0

Table 4.3: Frequency (in %) with which the correct MARX(0,1,1) model is selected

For $\varphi_1 \neq 0$, we see the same pattern in every case: identification of the correct model increases with β_1 and with sample size T . For a higher value of β_1 , the cross-covariance term becomes more important in determining the autocovariance of y_t , which is different for a purely causal and purely noncausal MARX. Because

MLE for the pure ARX case is established in Hannan et al. (1980). Estimation by OLS yields similar results.

of this, Gaussian MLE is able to distinguish between the two specifications in contrast to the case without exogenous regressors. In the same spirit, Cubadda et al. (2017) show that reduced rank restrictions help to identify purely causal from purely noncausal VAR models in a Gaussian framework whereas unrestricted models are not identifiable by Gaussian MLE.

4.4.3 Model Selection

Lanne and Saikkonen (2011a) propose a two-step approach to perform model selection for mixed causal-noncausal models $\text{MAR}(r, s)$. In a first step, purely causal autoregressive processes are estimated by OLS or Gaussian MLE and the lag order p is determined by conventional information criteria like AIC, BIC and HQ.¹⁴ As soon as p is fixed, one selects a model among all $\text{MAR}(r, s)$ with $p = r + s$. The model that attains the highest value for the log-likelihood at its estimated parameters is chosen to be the final model. This simulation study checks to what extent both steps are still valid in the MARX framework. To that end, we simulate (4.22) with $\phi_1 = 0.3$, $\varphi_1 = 0.5$ and $\beta_1 = 0.3$, $\varepsilon_t \stackrel{iid}{\sim} t(3, 0)$ and $x_{1,t} \stackrel{iid}{\sim} t(2, 0)$. Purely causal $\text{ARX}(p, 1)$ models are estimated by Gaussian MLE, where $p = 0, \dots, 4$. Table 4.4 shows the percentages with which AIC, BIC and HQ select a certain order p (true order equals 2). As comparison, the same exercise has been done on the MAR model. That is, we consider the same specification only without exogenous variable $x_{1,t}$, i.e. $\beta_1 = 0$. The corresponding frequencies for the MAR model can be found in Table 4.5.

p	$T = 100$			$T = 200$			$T = 500$			$T = 1000$		
	AIC	BIC	HQ	AIC	BIC	HQ	AIC	BIC	HQ	AIC	BIC	HQ
0	0.3	0.3	0.3	0.3	0.3	0.3	0.0	0.0	0.0	0.0	0.0	0.0
1	61.6	82.4	72.6	55.3	74.8	65.3	30.0	47.0	34.9	0.2	4.1	0.7
2	33.5	16.4	25.2	41.8	24.1	32.9	67.1	52.9	63.8	96.4	95.5	97.6
3	4.6	0.9	1.9	2.6	0.8	1.6	2.9	0.1	1.3	3.4	0.4	1.7
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 4.4: Frequency (in %) with which the autoregressive orders p and (r, s) are selected for the MARX model

We can see that all information criteria tend to underestimate the true lag order (especially BIC) in small samples. The performance improves when T grows

¹⁴In empirical work, it is advised to perform diagnostic tests to see whether additional lags are needed to remove autocorrelation from the series. Also a normality test on the residuals might be performed to test for signs of noncausality. A description of the model selection procedure can be found in Hecq et al. (2016).

p	$T = 100$			$T = 200$			$T = 500$			$T = 1000$		
	AIC	BIC	HQ	AIC	BIC	HQ	AIC	BIC	HQ	AIC	BIC	HQ
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	41.0	67.7	53.3	26.3	56.8	39.0	3.5	19.6	8.1	0.0	1.8	0.2
2	52.7	30.1	43.2	63.8	40.8	56.0	86.9	78.9	87.3	84.7	96.5	93.4
3	6.3	2.2	3.5	10.0	2.4	5.0	9.6	1.5	4.6	15.3	1.7	6.4
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 4.5: Frequency (in %) with which the autoregressive orders p and (r, s) are selected for the MAR model

larger but at $T = 500$ we still observe correct autoregressive order selection in only around 65% of the cases. Whereas it is well-known that information criteria are derived from asymptotic properties and thus might not perform optimally in finite samples (see e.g., Hurvich and Tsai, 1989), the performance in the MARX setup for $T \in \{100, 200, 500\}$ is considerably worse when compared to the same exercise for MAR models. This stresses the usage of diagnostic tests to discover model fit improvements. For instance, in the empirical section we also test for autocorrelation (Box-Pierce and LM tests) after having estimated pseudo causal models determined by information criteria. We will adapt the lag length (increase p) if necessary.

In a second exercise, we suppose the correct total autoregressive order $p = 2$ is known and investigate the selection of $\text{MARX}(r, s, 1)$ and $\text{MAR}(r, s)$ models with $r + s = 2$ based on the highest log-likelihood. In Table 4.6, we observe that the model selection procedure improves with sample size, finding the correct $\text{MARX}(1, 1, 1)$ specification in more than 80% of the cases for $T = 500$ and more than 90% when $T = 1000$. As datasets in empirical studies are often smaller, we advise practitioners to interpret the results with caution. We find that the correct model is only selected in a little more than half of the cases when $T = 100$, which suggests the use of complementary analysis (e.g., bootstrap or cross-validation criteria). In comparison, the MAR model selection is a lot more precise, as the correct model is already chosen in roughly 80% of the cases when $T = 100$.

	$T = 100$			$T = 200$			$T = 500$			$T = 1000$		
	(2,0)	(1,1)	(0,2)	(2,0)	(1,1)	(0,2)	(2,0)	(1,1)	(0,2)	(2,0)	(1,1)	(0,2)
MARX	35.5	54.4	10.1	29.0	68.8	2.2	16.9	83.0	0.1	6.8	93.2	0.0
MAR	3.3	78.7	18.0	0.2	95.4	4.4	0.0	99.9	0.1	0.0	100.0	0.0

Table 4.6: Lag-lead order (r, s) selected by highest log-likelihood with fixed $p = 2$ for both MARX and MAR model.

We find that the results of the model selection procedure are sensitive to the values of the parameters chosen in the DGP. Consider for example the case in which $\phi_1 = 0.1$ and $\varphi_1 = 0.7$. This will lead to an overselection of first-order models in the first step of the model selection procedure. Due to the fact that the noncausal parameter value is much larger than the causal one, a noncausal model will be selected in the second step. If we consider $\phi_1 = 0.5$ instead, the overselection of first-order models is less likely in the first step. Hence, it is also more probable that the “correct” mixed causal-noncausal model is eventually chosen in the second step.

4.5 Empirical Application

4.5.1 The Data

We consider non seasonally adjusted monthly commodity prices $CP_{i,t}$ from 1980:01 to 2016:10, i.e., 442 observations for $i = 1, \dots, 5$ indexes released by the IMF.¹⁵ These are benchmark prices which are representative of the global market. They are determined by the largest exporter of a given commodity. IMF releases many different individual commodity prices but we only focus on the following indexes for this study:

- BEVE: Beverage Price Index, 2005 = 100, includes Coffee, Tea, and Cocoa,
- INDU: Industrial Inputs Price Index, 2005 = 100, includes Agricultural Raw Materials and Metals Price Indices,
- RAWM: Agricultural Raw Materials Index, 2005 = 100, includes Timber, Cotton, Wool, Rubber, and Hides Price Indices,
- META Metals Price Index, 2005 = 100, includes Copper, Aluminum, Iron Ore, Tin, Nickel, Zinc, Lead, and Uranium Price Indices,
- OIL: Crude Oil (petroleum), Price index, 2005 = 100, simple average of three spot prices; Dated Brent, West Texas Intermediate, and the Dubai Fateh.

We also consider for the same period S_t , the trade weighted U.S. dollar index: broad, index Jan 1997=100, monthly, not seasonally adjusted as well as the level

¹⁵IMF Primary Commodity Prices, see <http://www.imf.org/external/np/res/commod/index.aspx>.

of the industrial production index (IP_t). These series are taken from the Federal Reserve Bank of St. Louis database.¹⁶ As commodities are mainly priced in dollar we can expect a contemporaneous negative relation between commodity prices and the U.S. exchange rate. When production increases, more input is needed and this has a positive effect on some of the commodities.

The way one has to detrend series before identifying MAR models is an ongoing debate. Hencic and Gouriéroux (2014) fit a cubic deterministic trend to bitcoin data. We decide to rely on usual unit root analysis. Using ADF tests, we do not reject a unit root at 5% significance level in each series. We consequently work with monthly growth rates $\Delta cp_{i,t} = (1 - L) \ln CP_{i,t}$, $\Delta ip_t = (1 - L) \ln IP_t$ and $\Delta s_t = (1 - L) \ln S_t$. These series are displayed in Figure 4.2.

4.5.2 From Expectation Models to MARX

Lanne and Luoto (2013) directly link mixed causal-noncausal models to the analysis of inflation using the hybrid NKPC. We can do the same with our relationship (4.1) in the introductory section but with two regressors (exchange rate and industrial production index)

$$\Delta cp_t = \beta_b \Delta cp_{t-1} + \beta_f \mathbb{E}(\Delta cp_{t+1} | \Omega_t) + \vartheta_1 \Delta s_t + \vartheta_2 \Delta ip_t + u_t. \quad (4.23)$$

By means of replacing expectations by the future realized growth rate of the commodity price plus a martingale difference ξ_{t+1} and putting the four parts ξ_{t+1} , Δs_t , Δip_t and u_t into the error term, one seemingly obtains an MAR(1,1) with a newly defined disturbance term, say η_t . However, since η_t contains the variable Δip_t and Δs_t , it is likely to be autocorrelated. Lanne and Luoto (2013) assume a MAR($r - 1, s - 1$) structure on η_t to show that the process for Δcp_t can be represented as a MAR(r, s). However, we are not certain that the regressors follow MAR($r - 1, s - 1$) dynamics and besides the direct effect of the fundamentals Δs_t and Δip_t on Δcp_t is lost in this mixed causal-noncausal model. For this reason, we prefer rearranging terms as follows

$$\Delta cp_t = \beta_b \Delta cp_{t-1} + \beta_f \Delta cp_{t+1} + \vartheta_1 \Delta s_t + \vartheta_2 \Delta ip_t + \underbrace{u_t + \beta_f \xi_{t+1}}_{\omega_{t+1}}. \quad (4.24)$$

¹⁶Series name TWEXBMTH and INDPROD at <https://fred.stlouisfed.org>.

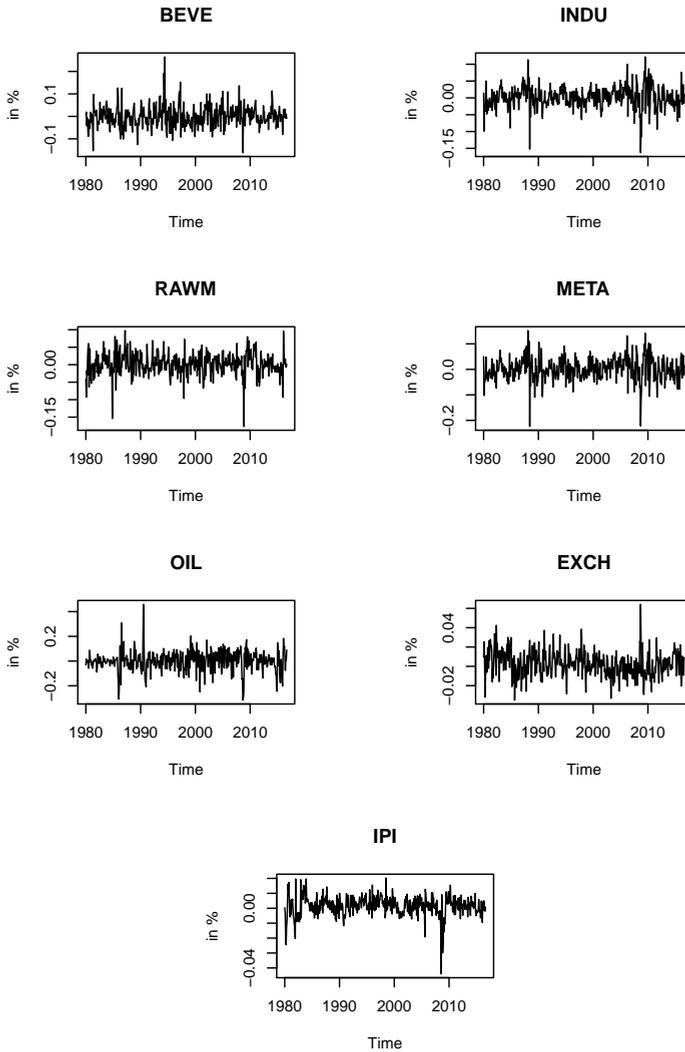


Figure 4.2: Growth rates of commodity prices, exchange rate and industrial production index

Now we lag (4.24) by one period and subsequently divide this equation by β_f to obtain

$$\beta_f^{-1} \Delta cp_{t-1} = \Delta cp_t + \beta_b \beta_f^{-1} \Delta cp_{t-2} + \vartheta_1 \beta_f^{-1} \Delta s_{t-1} + \vartheta_2 \beta_f^{-1} \Delta ip_{t-1} + \beta_f^{-1} \omega_{t+1},$$

which can be rewritten in the following way

$$(1 - \beta_f^{-1}L + \beta_b\beta_f^{-1}L^2)\Delta cp_t = -\vartheta_1\beta_f^{-1}\Delta s_{t-1} - \vartheta_2\beta_f^{-1}\Delta ip_{t-1} - \beta_f^{-1}\omega_t. \quad (4.25)$$

We want to write $(1 - \beta_f^{-1}L + \beta_b\beta_f^{-1}L^2)$ as $(1 - \phi L)(1 - \varphi^* L)$, where $|\phi| < 1$ and $|\varphi^*| > 1$. That is, we split the original polynomial in two different ones: one having all roots outside the unit circle $[\phi(z)]$ and one having its roots inside the unit circle $[\varphi^*(z)]$. With plausible values of β_b and β_f , this can be done by taking

$$\phi = \frac{1}{2} \left(\beta_f^{-1} - \sqrt{\beta_f^{-2} - 4\beta_f^{-1}\beta_b} \right) \quad \text{and} \quad \varphi^* = \frac{1}{2} \left(\beta_f^{-1} + \sqrt{\beta_f^{-2} - 4\beta_f^{-1}\beta_b} \right),$$

as was shown in Lanne and Luoto (2013). Following Lanne and Saikkonen (2011a), we can rewrite the polynomial with roots inside the unit circle as a polynomial in reverse time with roots outside the unit circle. That is,

$$\begin{aligned} (1 - \phi z)(1 - \varphi^* z) &= (1 - \phi z) \left[-\varphi^* z \left(-\frac{1}{\varphi^*} z^{-1} + 1 \right) \right] \\ &= -\varphi^* z (1 - \phi z) (1 - \varphi z^{-1}), \end{aligned}$$

where $\varphi = \frac{1}{\varphi^*}$. The polynomial in (4.25) can be replaced with this result to obtain

$$-\varphi^* L (1 - \phi L) (1 - \varphi L^{-1}) \Delta cp_t = -\vartheta_1 \beta_f^{-1} \Delta s_{t-1} - \vartheta_2 \beta_f^{-1} \Delta ip_{t-1} - \beta_f^{-1} \omega_t,$$

which by rearranging terms, reduces to a mixed causal-noncausal model, i.e.,

$$(1 - \phi L) (1 - \varphi L^{-1}) \Delta cp_t = \underbrace{\vartheta_1 (\varphi^* \beta_f)^{-1}}_{\beta_1} \Delta s_t + \underbrace{\vartheta_2 (\varphi^* \beta_f)^{-1}}_{\beta_2} \Delta ip_t + \underbrace{(\varphi^* \beta_f)^{-1} \omega_{t+1}}_{\varepsilon_t}.$$

We note that ε_t is an *iid* error term, as it is a function of ω_{t+1} which.

We consequently obtain the MARX(1,1,2) model. Note that we consider a simple one-lag one-lead example from the outset. Introducing more past and future expected terms and more regressors would yield higher order MARX(r, s, q) as solutions. Relaxing the *iid* assumption on ω_{t+1} and allowing it to have an autoregressive MAR($r - 1, s - 1$) structure leads to the same conclusion. This last proposal is, however, not in the spirit of rational expectations theory, which dictates that the realization will only deviate from the expectation because of information that could not have been known when the expectation was formed.

4.5.3 Identification of Pseudo ARDL models

The first step of our modelling strategy consists in fitting for each $\Delta cp_{i,t}$ series an OLS regression on an intercept, their own lags and the covariates Δs_t and Δip_t :

$$a(L)\Delta cp_{i,t} - b_1\Delta s_t - b_2\Delta ip_t - c = \epsilon_t. \quad (4.26)$$

The first part of Table 4.7 reports that step and provides the results concerning models chosen. In all cases, p_{max} is set to 8. The simulation results of Section 4.4.4 show that BIC tends to underestimate the true order more often than AIC and HQ. Since there is no clear evidence that one of these two performs better than the other, we decide to rely on HQ. However, we always check correlograms and perform LM tests for the null of no autocorrelation and add lags when it is necessary. We put that result, denoted $p(Q)$, in the second row of Table 4.7. It can be seen that only for the oil commodity series a departure from the result of HQ is necessary to get a white noise. Additionally, we start searching for autoregressive distributed lag models, i.e., pseudo ARDL($p_{max}, p_{max}, p_{max}$)¹⁷ to verify the inclusion of only the contemporaneous impact of Δs_t and Δip_t . Indeed, as we end up with ARDL($p_i, 0, 0$) for all commodity series, our choice is justified. This simple first step can be done using (for instance) EViews ARDL estimation features.¹⁸ Moreover, a simple regression shows that we do not reject the null of linear independence between the two exogenous variables.

The second part of Table 4.7 reports the estimation results (HCSE robust standard errors in brackets). We observe that commodity prices depend on their own lags as well as on the exchange rate and industrial production (except for BEVE). The highest negative effect of Δs_t is on the OIL index, a result that makes sense given that oil products heavily depend on exports and hence are more negatively influenced by an increase of the USD. The last row of Table 4.7 reports the value of the Jarque-Bera normality test. It is observed that we reject the null of normality in every equation and hence that we are able to discriminate components of the MARX model using non-Gaussian MLE.

¹⁷The first argument denotes the amount of lags of the dependent variable, the second and third argument the amount of lags for the exogenous variables.

¹⁸The same step can be performed in the R package **MARX** by allowing for lags in \mathbf{X}_t and comparing information criteria. This approach is however more cumbersome.

	Commodities				
	INDU	META	OIL	BEVE	RAWM
p_{HQ}	2	1	1	1	1
$p(Q)$	2	1	4	1	1
c	0.002 (0.0014)	0.002 (0.002)	0.003 (0.004)	0.001 (0.002)	0.001 (0.002)
a_1	0.185 (0.058)	0.187 (0.062)	0.298 (0.066)	0.280 (0.066)	0.183 (0.058)
a_2	0.091 (0.056)		-0.071 (0.055)		
a_3			0.027 (0.054)		
a_4			-0.122 (0.052)		
b_1	-0.889 (0.145)	-1.251 (0.201)	-1.673 (0.341)	-0.587 (0.176)	-0.453 (0.128)
b_2	0.699 (0.229)	1.048 (0.288)	1.152 (0.558)	0.467 (0.275)	0.604 (0.313)
\bar{R}^2	0.226	0.204	0.181	0.112	0.091
JB	101.33	99.82	132.47	107.64	143.38

Table 4.7: Estimation results - pseudo causal models

4.5.4 Identification and Estimation of MARX

Once the number of lags p_i ($i = 1, \dots, 5$) in ARDL models with the contemporaneous Δs_t and $\Delta i p_t$ are determined for each commodity price i , we estimate every possible MARX($r_i, s_i, 2$) models with $p_i = r_i + s_i$. We choose the model that gives the highest log-likelihood values. Table 4.8 reports the final results for each commodity. The values for ϕ (resp. φ) are the estimated coefficients of the lag (resp. lead) polynomials. We can observe some differences in the dynamics of the commodities. It emerges that we have purely causal autoregressive models for commodities INDU, META and RAWM. We have a purely noncausal specification for BEVE and a mixed model for OIL. Distributions are rather leptokurtic: the smallest value for the degrees of freedom parameter is $\hat{\nu} = 2.89$ for OIL, the largest value is $\hat{\nu} = 5.33$ for RAWM. The negative impact of exchange rate is more pronounced for OIL and the smallest value is for the raw material commodity index. The industrial production index is not significantly different from zero in the OIL and BEVE equations.

MARX(2,0,2)						
INDU	ϕ_1	0.247 (0.038)	β_1	-0.722 (0.089)	c	0.001 (0.002)
	ϕ_2	0.096 (0.038)	β_2	0.764 (0.175)	$[\nu, \sigma]$	[4.165, 0.021]
MARX(1,0,2)						
META	ϕ_1	0.265 (0.038)	β_1	-1.062 (0.131)	c	0.002 (0.002)
			β_2	0.992 (0.255)	$[\nu, \sigma]$	[5.299, 0.032]
MARX(1,0,2)						
RAWM	ϕ_1	0.191 (0.041)	β_1	-0.374 (0.098)	c	0.001 (0.001)
			β_2	0.457 (0.191)	$[\nu, \sigma]$	[5.338, 0.024]
MARX(2,2,2)						
OIL	ϕ_1	-0.491 (0.036)	β_1	-1.100 (0.212)	c	0.007 (0.003)
	ϕ_2	-0.176 (0.036)	β_2	0.284 (0.413)	$[\nu, \sigma]$	[2.890, 0.048]
	φ_1	0.725 (0.034)				
	φ_2	-0.241 (0.034)				
MARX(0,1,2)						
BEVE	φ_1	0.295 (0.039)	β_1	-0.409 (0.141)	c	-0.001 (0.001)
			β_2	0.241 (0.276)	$[\nu, \sigma]$	[4.954, 0.034]

Table 4.8: Estimation results - MARX models

Since Lanne and Saikkonen (2011a) claim that the errors in mixed causal-noncausal models contain effects of omitted variables that are predictable by the considered series, the necessity of exogenous variables in a noncausal model could be questioned. Significance of the exchange rate in every series and the production index in three commodities in this empirical application seems to indicate that not all of its effect is predictable by the respective commodity price dynamics. Indeed, estimating $MAR(r, s)$, namely when excluding the exogenous regressors, we obtain $MAR(1,1)$ for INDU, $MAR(0,1)$ for BEVE, $MAR(1,0)$ for META, $MAR(0,4)$ for OIL and $MAR(0,1)$ for RAWM. This leads to having a noncausal component in each series but META and illustrates that noncausal models can indeed capture the information that economic agents have but not the econometrician. Explicitly adding regressors gives different models and hence justifies the use of the $MARX(r, s, q)$ models both for forecasting and for the understanding of economic relationships. Lastly, we can also illustrate the identification feature raised in Section 4.2.2. Let us consider the META commodity as an example, as we obtained a purely causal model and coefficients for both exogenous regressors which are high and significantly different from zero. If we now estimate a purely causal $MARX(1,0,2)$ and purely noncausal $MARX(0,1,2)$ by OLS, we obtain values for

the log-likelihood of respectively 788.65 and 785.59, which indicate that the causal pattern is favored.

4.6 Conclusion

This chapter proposes to estimate mixed causal-noncausal models by non-Gaussian MLE when additional regressors are present. We have in mind the estimation of structural relationships subject to expectation schemes such as the new Hybrid Phillips curve or lag-augmented present value models. Many empirical macroeconomic equations are covered by this framework. We provide a successful empirical illustration on the relation between commodity prices, the exchange rate and the industrial production index.

The one-step approach to estimating MARX is easy to implement and the estimation of the standard errors that we propose is quite robust to computational overflows. It allows to estimate directly the impact of exogenous variables without the need to augment the MAR with leads and lags (and to lose the impact of \mathbf{X}_t) or to use a two-step approach as in Lof and Nyberg (2017). In theory, we find that, the presence of exogenous regressors have the appealing feature that one can discriminate between mixed causal-noncausal specifications in a Gaussian framework.

4.A From Transfer Function Model to MARX

For expository purposes, we consider a single explanatory variable denoted x_t^* . The transfer function model is given by

$$y_t = \psi^*(L)x_t^* + n_t, \quad (4.27)$$

where n_t is a noise process assumed to follow a stationary AR process, $a(L)n_t = \varepsilon_t^*$. The ARX (or ARDL) model can be motivated from (4.27) by assuming that the transfer function operator can be expressed in a rational factorization as $\psi^*(z) = a(z)^{-1}\theta^*(z)$. Multiplying (4.27) by $a(L)$ yields

$$\begin{aligned} a(L)y_t &= \theta^*(L)x_t^* + a(L)n_t \\ &= \theta^*(L)x_t^* + \varepsilon_t^*, \end{aligned} \quad (4.28)$$

which is the usual ARX(p, k) model representation when $\deg(a(z)) = p$ and $\deg(\theta^*(z)) = k$. If all roots of $a(z)$ lie outside the unit circle, the process is stationary which implies that estimation and inference can directly be conducted. Breidt et al. (1991) consider the more complex case in which r roots lie outside the unit circle and s inside ($r + s = p$) and propose to factorize the polynomial to obtain

$$\phi(L)\varphi^*(L)y_t = \theta^*(L)x_t^* + \varepsilon_t^*.$$

Lanne and Saikkonen (2011a) propose to rewrite $\varphi^*(z)$ in terms of the lead operator and obtain the relation $\varphi(z^{-1}) = -\varphi_s^*z^s\varphi^*(z)$. By rearranging terms, we find

$$\begin{aligned} \phi(L)\varphi(L^{-1})y_t &= \left(-\frac{1}{\varphi_s^*} + \frac{\theta_1^*}{\varphi_s^*} + \dots + \frac{\theta_k^*}{\varphi_s^*} \right) x_{t+s}^* - \frac{1}{\varphi_s^*} \varepsilon_{t+s}^* \\ &= \theta(L)x_t + \varepsilon_t. \end{aligned} \quad (4.29)$$

In case only a contemporaneous value of x_t enters the system, take $\psi^*(z) = a(z)^{-1}\beta$. Note that the derivation can easily be extended to q regressors by defining $\psi^*(z) = [\psi_1^*(z), \dots, \psi_q^*(z)]'$ and considering X_t^* . In the distributed lag case take $\psi^*(z) = a(z)^{-1}\theta^*(z)$ with $\theta^*(z) = [\theta_1^*(z), \dots, \theta_k^*(z)]'$; in the contemporaneous case define $\psi^*(z) = a(z)^{-1}\beta$ with $\beta \in \mathbb{R}^q$. In case one wants to allow for (mixed) dynamics in the exogenous regressors, it seems more natural to model such a process

as a VAR. The mixed VAR model (see e.g., Lanne and Saikkonen, 2013; Davis and Song, 2012) accommodates this structure.

4.B Approximate Likelihood Function

Define $\mathbf{b} = \beta' \tilde{\mathbf{x}}$ such that $\mathbf{z} = \mathbf{B}\mathbf{A}\mathbf{y} - \beta' \tilde{\mathbf{x}} = \mathbf{B}\mathbf{A}\mathbf{y} - \mathbf{b}$ (See case 5 in the proof of Lemma 4.1 in Appendix 4.C). Assume \mathbf{B} and \mathbf{A} are invertible. We are interested in the inverse transformation, i.e. $\mathbf{y} = \mathbf{Q}(\mathbf{z} + \mathbf{b})$, where $\mathbf{Q} = \mathbf{B}^{-1}\mathbf{A}^{-1}$. Let \mathbf{Q} be a (2×2) matrix, then we have

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} q_1 & q_2 \\ q_3 & q_4 \end{bmatrix} \left(\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \right), \quad (4.30)$$

with the following functions

$$y_1 = g_1(z_1, z_2) = q_1 z_1 + q_2 z_2 + b_1, \quad (4.31)$$

$$y_2 = g_2(z_1, z_2) = q_3 z_1 + q_4 z_2 + b_2. \quad (4.32)$$

The Jacobian is given as the matrix of all partial derivatives from \mathbf{y} to \mathbf{z} , i.e.

$$\mathbf{J} = \begin{bmatrix} \frac{\partial y_1}{\partial z_1} & \frac{\partial y_1}{\partial z_2} \\ \frac{\partial y_2}{\partial z_1} & \frac{\partial y_2}{\partial z_2} \end{bmatrix} = \begin{bmatrix} q_1 & q_2 \\ q_3 & q_4 \end{bmatrix} = \mathbf{Q}. \quad (4.33)$$

Then the joint density of y_1 and y_2 is given as:

$$f_{y_1, y_2}(y_1, y_2) = \frac{1}{|\det(\mathbf{Q})|} f_{z_1, z_2}(\mathbf{B}\mathbf{A}\mathbf{y} - \mathbf{b}) \quad (4.34)$$

It is well-known that this result can be generalized to higher orders (e.g., Casella and Berger, 2002, p. 185). From Proposition 4.1 we know that the information sets (i) and (vi) are observationally equivalent. Using the transformations in (4.36)-(4.37) and $\mathbf{Q} = \mathbf{B}^{-1}\mathbf{A}^{-1}$, we find that the probability density of the process y_t can be represented in the following way:

$$\begin{aligned} f_{y; \lambda}(\mathbf{y}) &= \frac{1}{|\det(\mathbf{Q})|} f_z(\mathbf{B}\mathbf{A}\mathbf{y} - \mathbf{b}; \lambda) \\ &= |\det(\mathbf{A})| |\det(\mathbf{B})| h_V(\mathbf{B}\mathbf{A}\mathbf{y} - \mathbf{b}) f_\varepsilon(\mathbf{B}\mathbf{A}\mathbf{y} - \mathbf{b}; \lambda) h_U(\mathbf{B}\mathbf{A}\mathbf{y} - \mathbf{b}) \end{aligned}$$

$$\begin{aligned}
 &= |\det(\mathbf{A})| h_V(\mathbf{B}\mathbf{A}\mathbf{y} - \mathbf{b}) \left(\prod_{t=r+1}^{T-s} f_\sigma(\mathbf{B}\mathbf{A}\mathbf{y} - \mathbf{b}; \boldsymbol{\lambda}) \right) h_U(\mathbf{B}\mathbf{A}\mathbf{y} - \mathbf{b}) \\
 &= h_V(\varphi(L^{-1})y_1, \dots, \varphi(L^{-1})y_r) \left(\prod_{t=r+1}^{T-s} f_\sigma(\phi(L)\varphi(L^{-1})y_t - \boldsymbol{\beta}'\mathbf{X}_t; \boldsymbol{\lambda}) \right) \\
 &\quad h_U(\phi(L)y_{T-s+1}, \dots, \phi(L)y_T) |\det(\mathbf{A})|, \tag{4.35}
 \end{aligned}$$

where \mathbf{A} and \mathbf{B} are two nonsingular matrices with $\det(\mathbf{B}) = 1$; h_V and h_U are the joint densities of (v_1, \dots, v_r) and (u_{T-s+1}, \dots, u_T) respectively. Independence of the blocks (v_1, \dots, v_r) , $(\varepsilon_{r+1}, \dots, \varepsilon_{T-s})$ and (u_{T-s+1}, \dots, u_T) is applied in the second equality and the definition of the filtered values as presented in (4.8) and (4.9) in the fourth equality. Since $\det(\mathbf{A})$ is independent of sample size, the density of y_t can be approximated by the second term in (4.35).

4.C Proofs

Proof of Proposition 4.1

Let \sim denote equivalence in information sets. To show that (i), (ii) and (iii) are equivalent is similar to showing that (i'), (ii') and (iii') are equivalent. We prove (i') \sim (ii'), (i') \sim (iii'), (ii) \sim (iv), (iii) \sim (v) and (i) \sim (vi).

Case 1: (i') \sim (ii')

Note that by the definition of u in equation (4.8), (ii') $(y_1, \dots, y_r, u_{r+1}, \dots, u_T) = (y_1, \dots, y_r, \phi(L)y_{r+1}, \dots, \phi(L)y_T)$. Since $u_{r+1} = y_{r+1} - \phi_1 y_r - \dots - \phi_r y_1$ with y_1, \dots, y_r and u_{r+1} known, y_{r+1} is known. The same reasoning can be recursively applied to u_{r+2} up to u_T , leading to the desired result.

Case 2: (i') \sim (iii')

Note that by the definition of v in equation (4.9), (iii') $(v_1, \dots, v_{T-s}, y_{T-s+1}, \dots, y_T) = (\varphi(L^{-1})y_1, \dots, \varphi(L^{-1})y_{T-s}, y_{T-s+1}, \dots, y_T)$. Since $v_{T-s} = y_{T-s} - \varphi_1 y_{T-s+1} - \dots - \varphi_s y_T$ with y_{T-s+1}, \dots, y_T and v_{T-s} known, y_{T-s} is known. The same reasoning can be recursively applied to v_{T-s-1} up to v_1 , leading to the desired result.

Hence, since (i'), (ii') and (iii') are equivalent, we know that (i), (ii) and (iii) are as well (as all information sets are augmented with the same information).

Case 3: (ii) \sim (iv)

Note that (iv) $(y_1, \dots, y_r, \varepsilon_{r+1}, \dots, \varepsilon_{T-s}, u_{T-s+1}, \dots, u_T) = (y_1, \dots, y_r, \varphi(L^{-1})u_{r+1} - \beta' \mathbf{X}_{r+1}, \dots, \varphi(L^{-1})u_{T-s} - \beta' \mathbf{X}_{T-s}, u_{T-s+1}, \dots, u_T)$ by using the second equality in equation (4.8). Since $\varepsilon_{T-s} = u_{T-s} - \varphi_1 u_{T-s+1} - \dots - \varphi_s u_T - \beta' \mathbf{X}_{T-s}$ with $u_{T-s+1}, \dots, u_T, \mathbf{X}_{T-s}$ and ε_{T-s} known, u_{T-s} is known. The same reasoning can be recursively applied to u_{T-s-1} up to u_{r+1} , leading to the desired result.

Case 4: (iii) \sim (v)

Note that (v) $(v_1, \dots, v_r, \varepsilon_{r+1}, \dots, \varepsilon_{T-s}, y_{T-s+1}, \dots, y_T) = (v_1, \dots, v_r, \phi(L)v_{r+1} - \beta' \mathbf{X}_{r+1}, \dots, \phi(L)v_{T-s} - \beta' \mathbf{X}_{T-s}, y_{T-s+1}, \dots, y_T)$ by using the second equality in equation (4.9). Since $\varepsilon_{r+1} = v_{r+1} - \phi_1 v_r - \dots - \phi_r v_1 - \beta' \mathbf{X}_{r+1}$ with $v_1, \dots, v_r, \mathbf{X}_{r+1}$ and ε_{r+1} known, v_{r+1} is known. The same reasoning can be recursively applied to v_{r+2} up to v_{T-s} , leading to the desired result.

Case 5: (i) \sim (vi)

To show: $(y_1, \dots, y_T, \mathbf{X}_r, \dots, \mathbf{X}_{T-s}) \sim (v_1, \dots, v_r, \varepsilon_{r+1}, \dots, \varepsilon_{T-s}, u_{T-s+1}, \dots, u_T)$. Denote the vector corresponding to the first information set by $\tilde{\mathbf{y}}$ and the second one by \mathbf{z} . This statement can be proven using the algebra in Lanne and Saikkonen (2011a). Define the vectors $\mathbf{w} = [v_1, \dots, v_{T-s}, u_{T-s+1}, \dots, u_T]'$ and $\mathbf{y} = [y_1, \dots, y_T]'$. Then

$$\begin{bmatrix} v_1 \\ \vdots \\ v_{T-s} \\ u_{T-s+1} \\ \vdots \\ v_T \end{bmatrix} = \begin{bmatrix} y_1 - \varphi_1 y_2 - \dots - \varphi_s y_{s+1} \\ \vdots \\ y_{T-s} - \varphi_1 y_{T-s+1} - \dots - \varphi_s y_T \\ y_{T-s+1} - \phi_1 y_{T-s} - \dots - \phi_r y_{T-s+1-r} \\ \vdots \\ y_T - \phi_1 y_{T-1} - \dots - \phi_r y_{T-r} \end{bmatrix} = \mathbf{A} \begin{bmatrix} y_1 \\ \vdots \\ y_{T-s} \\ y_{T-s+1} \\ \vdots \\ y_T \end{bmatrix}, \quad (4.36)$$

which can be written as $\mathbf{w} = \mathbf{A}\mathbf{y}$. Now define $\tilde{\mathbf{x}} = [\underbrace{0, \dots, 0}_{r \text{ times}}, \mathbf{X}_{r+1}, \dots, \mathbf{X}_{T-s}, \underbrace{0, \dots, 0}_{s \text{ times}}]'$.

Similarly, we can form the following system of equations (with slight abuse of notation, as \mathbf{X}_t is a vector) for the vector \mathbf{z} . That is, equation (4.37) on the next page shows that $\mathbf{z} = \mathbf{B}\mathbf{w} - \beta'\tilde{\mathbf{x}}$. Combining both systems of equations, we find that the vectors \mathbf{z} and $\tilde{\mathbf{y}}$ are related in the following way: $\mathbf{z} = \mathbf{B}\mathbf{A}\mathbf{y} - \beta'\tilde{\mathbf{x}}$, where \mathbf{y} and $\tilde{\mathbf{x}}$ combined form the information set $\tilde{\mathbf{y}}$. Since the matrices \mathbf{B} and \mathbf{A} as

well as the parameter vector β only contain the known parameters, this shows that these information sets are equivalent. Combining all cases shows that information sets (i) – (vi) are equivalent.

$$\begin{bmatrix} v_1 \\ \vdots \\ v_r \\ v_{r+1} - \phi_1 v_r - \dots - \phi_r v_1 - \beta' \mathbf{X}_{r+1} \\ \vdots \\ v_{T-s} - \phi_1 v_{T-s-1} - \dots - \phi_r v_{T-s-r} - \beta' \mathbf{X}_{T-s} \\ u_{T-s+1} \\ \vdots \\ u_T \end{bmatrix} = \mathbf{B} \begin{bmatrix} v_1 \\ \vdots \\ v_r \\ v_{r+1} \\ \vdots \\ v_{T-s} \\ u_{T-s+1} \\ \vdots \\ u_T \end{bmatrix} - \beta' \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \mathbf{X}_{r+1} \\ \vdots \\ \mathbf{X}_{T-s} \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{4.37}$$

Proof of Lemma 4.2

For the proof of this lemma and the next theorem, we need some additional notation. Define $e_t = \frac{f'_\sigma(\varepsilon_t; \lambda)}{f_\sigma(\varepsilon_t; \lambda)} = \frac{f'(\varepsilon_t/\sigma; \lambda)}{\sigma f(\varepsilon_t/\sigma; \lambda)}$, $\tilde{\mathcal{J}} = \sigma^{-2} \mathcal{J}$, $\tilde{\mathcal{I}} = \sigma^{-2} \mathcal{I}$ and $n \equiv (T - p)$. Furthermore, let $x = \varepsilon_t/\sigma$, then we have that

$$\begin{aligned} \mathbb{E}(e_t^2) &= \mathbb{E} \left[\left(\frac{f'_\sigma(\varepsilon_t; \lambda)}{f_\sigma(\varepsilon_t; \lambda)} \right)^2 \right] \\ &= \int \left(\frac{f'_\sigma(\varepsilon_t; \lambda)}{f_\sigma(\varepsilon_t; \lambda)} \right)^2 f_\sigma(\varepsilon_t; \lambda) d\varepsilon_t \\ &= \sigma^{-2} \int \frac{(f'(x; \lambda))^2}{f(x; \lambda)} dx = \tilde{\mathcal{J}}, \end{aligned}$$

where we used the definitions of the density and \mathcal{J} . Also we have that

$$\begin{aligned} \mathbb{E}(e_t) &= \mathbb{E} \left(\frac{f'_\sigma(\varepsilon_t; \lambda)}{f_\sigma(\varepsilon_t; \lambda)} \right) \\ &= \int f'_\sigma(\varepsilon_t; \lambda) d\varepsilon_t \\ &= \sigma^{-1} f(x)|_{-\infty}^{\infty} = 0, \end{aligned}$$

which follows by the definition of the density and assumption (A3) in Breidt et al. (1991). To simplify future computations, we begin by noting that

$$\mathbb{E}(z_s e_t) = \begin{cases} 0, & \text{if } s \neq t, \\ -1, & \text{if } s = t, \end{cases} \quad (4.38)$$

which follows from the assumptions on the density and strict exogeneity between all exogeneous regressors and the error term. Now, for $i = 1, \dots, r$, we can show that

$$\begin{aligned} \mathbb{E} \left(\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \phi_i} \right) &= \mathbb{E}(-e_t v_{t-i}) \\ &= -\mathbb{E} \left(e_t \sum_{j=0}^{\infty} \alpha_j z_{t-i-j} \right) \\ &= 0. \end{aligned}$$

Hence, we note that \mathbf{V}_{t-1} and e_t are still independent as in Lanne and Saikkonen (2011a). Consequently, we still find

$$\begin{aligned} \text{Cov} \left(\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \phi} \right) &= \text{Cov}(-\mathbf{V}_{t-1} e_t) \\ &= \mathbb{E}(e_t^2) \mathbb{E}(\mathbf{V}_{t-1} \mathbf{V}_{t-1}') \\ &= \tilde{\mathcal{J}} \boldsymbol{\Gamma}_V, \end{aligned}$$

where $\boldsymbol{\Gamma}_V$ denotes the autocovariance matrix of the vector \mathbf{V}_{t-1} . Because $\mathbf{V}_{t-1} e_t$ is uncorrelated, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Cov} \left(\sum_{t=r+1}^{T-s} \frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \phi} \right) = \tilde{\mathcal{J}} \boldsymbol{\Gamma}_V.$$

Symmetrically, by using similar arguments, we can show for $i = 1, \dots, s$ that

$$\begin{aligned} \mathbb{E} \left(\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \varphi_i} \right) &= \mathbb{E}(-e_t u_{t+i}) \\ &= -\mathbb{E} \left(e_t \sum_{j=0}^{\infty} \delta_j z_{t+i+j} \right) \\ &= 0. \end{aligned}$$

That is, the independence of e_t and \mathbf{U}_{t+1} is preserved through strict exogeneity. Letting $\mathbf{\Gamma}_U$ be the autocovariance matrix of \mathbf{U}_{t+1} , we find that

$$\begin{aligned} \text{Cov}\left(\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}}\right) &= \text{Cov}(-\mathbf{U}_{t+1}e_t) \\ &= \mathbb{E}(e_t^2)\mathbb{E}(\mathbf{U}_{t+1}\mathbf{U}'_{t+1}) \\ &= \tilde{\mathcal{J}}\mathbf{\Gamma}_U. \end{aligned}$$

Because $\mathbf{U}_{t+1}e_t$ is uncorrelated, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Cov}\left(\sum_{t=r+1}^{T-s} \frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}}\right) = \tilde{\mathcal{J}}\mathbf{\Gamma}_U.$$

Lastly, we can apply the same logic for the parameter vector $\boldsymbol{\beta}$. Since for $i = 1, \dots, q$, we have that

$$\mathbb{E}\left(\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \beta_i}\right) = 0$$

by the independence of $x_{i,t}$ and ε_t . If we denote by $\mathbf{\Gamma}_X$, the autocovariance matrix of \mathbf{X}_t , it follows that

$$\begin{aligned} \text{Cov}\left(\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}}\right) &= \text{Cov}(-\mathbf{X}_t e_t) \\ &= \mathbb{E}(e_t^2)\mathbb{E}(\mathbf{X}_t\mathbf{X}'_t) \\ &= \tilde{\mathcal{J}}\mathbf{\Gamma}_X. \end{aligned}$$

Because $\mathbf{X}_t e_t$ is uncorrelated, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Cov}\left(\sum_{t=r+1}^{T-s} \frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}}\right) = \tilde{\mathcal{J}}\mathbf{\Gamma}_X.$$

We now characterize the covariances of the partials. To that end, we first notice that

$$\text{Cov}(z_{t-i}e_t, z_{k-j}e_k) = \begin{cases} \mathcal{I} + \tilde{\mathcal{J}} \sum_{m=1}^q \beta_m^2 \sigma_{x_m}^2 & \text{if } t = k, i = j = 0, \\ \mathcal{J} + \tilde{\mathcal{J}} \sum_{m=1}^q \beta_m^2 \sigma_{x_m}^2 & \text{if } t = k, i = j \neq 0, \\ 1 & \text{if } t \neq k, i = t - k, j = k - t, \\ 0 & \text{otherwise.} \end{cases} \quad (4.39)$$

Hence, using (4.38)-(4.39), we find that

$$\begin{aligned} \text{Cov}\left(\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \phi_i}, \frac{\partial g_k(\boldsymbol{\theta}_0)}{\partial \phi_j}\right) &= \begin{cases} \gamma_V(i-j)\tilde{\mathcal{J}}, & \text{if } t = k, 1 \leq i \leq j \leq r, \\ 0, & \text{otherwise.} \end{cases} \\ \text{Cov}\left(\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \varphi_i}, \frac{\partial g_k(\boldsymbol{\theta}_0)}{\partial \varphi_j}\right) &= \begin{cases} \gamma_U(i-j)\tilde{\mathcal{J}}, & \text{if } t = k, 1 \leq i \leq j \leq s, \\ 0, & \text{otherwise.} \end{cases} \\ \text{Cov}\left(\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \beta_i}, \frac{\partial g_k(\boldsymbol{\theta}_0)}{\partial \beta_j}\right) &= \begin{cases} \sigma_{x_i}^2 \tilde{\mathcal{J}}, & \text{if } t = k, i = j \neq 0, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Define $Q_m(i, j, a) \equiv \sum_{b=0}^{\infty} \delta_b \gamma_{x_m}(i + j + a + b)$. For the covariance matrix between $\partial g_t(\boldsymbol{\theta}_0)/\partial \phi$ and $\partial g_t(\boldsymbol{\theta}_0)/\partial \varphi$, first consider for $1 \leq i \leq r, 1 \leq j \leq s$:

$$\begin{aligned} \text{Cov}\left(\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \phi_i}, \frac{\partial g_k(\boldsymbol{\theta}_0)}{\partial \varphi_j}\right) &= \text{Cov}\left(\sum_{a=0}^{\infty} \alpha_a z_{t-i-a} e_t, \sum_{b=0}^{\infty} \delta_b z_{k+j+b} e_k\right) \\ &= \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} \alpha_a \delta_b \text{Cov}(z_{t-i-a} e_t, z_{k+j+b} e_k) \\ &= \begin{cases} \alpha_{t-k-i} \delta_{t-k-j}, & \text{for } t > k, \\ \tilde{\mathcal{J}} \sum_{a=0}^{\infty} \alpha_a \sum_{m=1}^q \beta_m^2 Q_m(i, j, a) & \text{for } t = k, \\ 0 & \text{for } t < k. \end{cases} \end{aligned}$$

The element (i, j) of the matrix $\frac{1}{n} \text{Cov}(\partial L_T(\boldsymbol{\theta}_0)/\partial \phi, \partial L_T(\boldsymbol{\theta}_0)/\partial \varphi)$ is

$$n \text{Cov}\left(\frac{1}{n} \sum_{t=r+1}^{T-s} \frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \phi_i}, \frac{1}{n} \sum_{k=r+1}^{T-s} \frac{\partial g_k(\boldsymbol{\theta}_0)}{\partial \varphi_j}\right)$$

$$\begin{aligned}
 &= \frac{1}{n} \sum_{t=r+1}^{T-s} \sum_{k=r+1}^{T-s} \text{Cov} \left(\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \phi_i}, \frac{\partial g_k(\boldsymbol{\theta}_0)}{\partial \varphi_j} \right) \\
 &= \frac{1}{n} \sum_{t=r+1}^{T-s} \sum_{k=r+1}^{T-s} \text{Cov}(-v_{t-i}e_t, -u_{k+j}e_k) \\
 &= \frac{1}{n} \sum_{t=r+1}^{T-s} \sum_{k=r+1}^{T-s} \left(\mathbb{1}_{\{t>k\}} \alpha_{t-k-i} \delta_{t-k-j} + \mathbb{1}_{\{t=k\}} \tilde{\mathcal{J}} \sum_{a=0}^{\infty} \alpha_a \sum_{m=1}^q \beta_m^2 Q_m(i, j, a) \right) \\
 &= \frac{1}{n} \left(\sum_{k=r+1}^{T-s-1} \sum_{t=k+1}^{T-s} \alpha_{t-k-i} \delta_{t-k-j} + n \tilde{\mathcal{J}} \sum_{a=0}^{\infty} \alpha_a \sum_{m=1}^q \beta_m^2 Q_m(i, j, a) \right) \\
 &= \frac{1}{n} \left(\sum_{k=r+1}^{T-s-1} \sum_{t=0}^{T-s-k-i} \alpha_t \delta_{t+i-j} \right) + \tilde{\mathcal{J}} \sum_{a=0}^{\infty} \alpha_a \sum_{m=1}^q \beta_m^2 Q_m(i, j, a) \\
 &\rightarrow \sum_{t=0}^{\infty} \alpha_t \delta_{t+i-j} + \tilde{\mathcal{J}} \sum_{a=0}^{\infty} \alpha_a \sum_{m=1}^q \beta_m^2 Q_m(i, j, a),
 \end{aligned}$$

where convergence of the first term follows from the geometric decay of the sequences $\{\alpha_t\}$ and $\{\delta_t\}$. Note that $\delta_{t+i-j} = 0$ for $t+i-j < 0$. The equalities follow from results presented earlier, the change of summands follows from imposing $t > k$.

Next, we consider the covariance between the partial derivatives of the log-likelihood with respect to the causal autoregressive parameters $\boldsymbol{\phi}$ and the parameter vector of the exogenous variables $\boldsymbol{\beta}$. That is,

$$\begin{aligned}
 \text{Cov} \left(\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \phi_i}, \frac{\partial g_k(\boldsymbol{\theta}_0)}{\partial \beta_j} \right) &= \text{Cov}(-v_{t-i}e_t, -x_{j,k}e_k) \\
 &= \mathbb{E} \left(\sum_{a=0}^{\infty} \alpha_a \varepsilon_{t-i-a} x_{j,k} e_k e_t \right) \\
 &\quad + \mathbb{E} \left(\sum_{a=0}^{\infty} \alpha_a \sum_{m=1}^q \beta_m x_{m,t-i-a} x_{j,k} e_k e_t \right) \\
 &= \begin{cases} \beta_j \tilde{\mathcal{J}} \sum_{a=0}^{\infty} \alpha_a \gamma_{x_j}(i+a) & \text{for } t=k, 1 \leq i \leq r, \\ 0 & \text{for } t \neq k, 1 \leq i \leq r. \end{cases}
 \end{aligned}$$

This outcome is independent of time t . Symmetrically, we compute the covariance between the partial derivatives of the log-likelihood with respect to the noncausal autoregressive parameters $\boldsymbol{\varphi}$ and the parameter vector of the exogenous variables $\boldsymbol{\beta}$:

$$\begin{aligned}
\text{Cov}\left(\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \varphi_i}, \frac{\partial g_k(\boldsymbol{\theta}_0)}{\partial \beta_j}\right) &= \text{Cov}(-u_{t+i}e_t, -x_{j,k}e_k) \\
&= \mathbb{E}\left(\sum_{b=0}^{\infty} \delta_b \varepsilon_{t+i+b} x_{j,k} e_k e_t\right) \\
&+ \mathbb{E}\left(\sum_{b=0}^{\infty} \delta_b \sum_{m=1}^q \beta_m x_{m,t+i+b} x_{j,k} e_k e_t\right) \\
&= \begin{cases} \beta_j \tilde{\mathcal{J}} \sum_{b=0}^{\infty} \delta_b \gamma_{x_j}(i+b) & \text{for } t = k, 1 \leq i \leq s, \\ 0 & \text{for } t \neq k, 1 \leq i \leq s. \end{cases}
\end{aligned}$$

The proof of asymptotic normality is similar to Breidt et al. (1991) and Lanne and Saikkonen (2011a). Define $\mathbf{M} = \text{diag}(\boldsymbol{\Sigma}, \boldsymbol{\Omega})$, $\mathbf{W}_t = \frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$ and note that $n \equiv (T - p)$. By the Cramér-Wold theorem, it suffices to show that for any vector \mathbf{a} of appropriate size,

$$\frac{1}{\sqrt{n}} \sum_{t=r+1}^{T-s} \mathbf{a}' \mathbf{W}_t \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{a}' \mathbf{M} \mathbf{a}). \quad (4.40)$$

Define the sequence of $(p + q + d + 1)$ dimensional random vectors $\{\mathbf{W}_{tm}, t \in \mathbb{Z}\}$ to be the partials defined in Section 4.3.1, where v_t , u_t and all $x_{i,t}$ for $i = 1, \dots, q$ are replaced by their representation in (4.11)-(4.12) and assumption (A2) with the sums truncated at a large positive integer m , i.e.,

$$v_t^{(m)} = \sum_{j=0}^m \alpha_j z_{t-j}, \quad u_t^{(m)} = \sum_{j=0}^m \delta_j z_{t+j} \quad \text{and} \quad x_{i,t}^{(m)} = c_i + \sum_{j=-m}^m \rho_{i,j} \eta_{i,t-j}.$$

It can be verified that $\mathbb{E}(\mathbf{W}_t) = \mathbf{0}$ and $\gamma_{\mathbf{W}_t}(0) + 2 \sum_{j=1}^{\infty} \gamma_{\mathbf{W}_t}(j) \neq \mathbf{0}$. This result also holds for \mathbf{W}_t replaced by \mathbf{W}_{tm} . Let \mathbf{M}_m be the matrix corresponding to \mathbf{M} , obtained by truncating u_t , v_t and \mathbf{X}_t . Then the stationary sequence $\{\mathbf{W}_{tm}, t \in \mathbb{Z}\}$ is $\max\{m+p, 2m\}$ dependent.¹⁹ Now that we verified the conditions, we can apply Theorem 6.4.2 in Brockwell and Davis (1991) to obtain

$$\frac{1}{\sqrt{n}} \sum_{t=r+1}^{T-s} \mathbf{a}' \mathbf{W}_{tm} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{a}' \mathbf{M}_m \mathbf{a}).$$

¹⁹The $m+p$ follows from writing u_t and v_t in their truncated representation, $2m$ follows from the processes in \mathbf{X}_t which have a two-sided MA representation truncated by m at both sides.

Now, it follows that for $m \rightarrow \infty$, $\mathbf{W}_{tm} \rightarrow \mathbf{W}_t$ (by definition) and thus $\mathbf{M}_m \rightarrow \mathbf{M}$. Because

$$\lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \text{Var} \left(\frac{1}{\sqrt{n}} \sum_{t=r+1}^{T-s} (\mathbf{a}' \mathbf{W}_{tm} - \mathbf{a}' \mathbf{W}_t) \right) = 0,$$

the convergence in (4.40) is immediate from Proposition 6.3.9 in Brockwell and Davis (1991). The positive definiteness of $\mathbf{\Sigma}$ can be established similar to the proof in Breidt et al. (1991). In the MARX case, the block matrix $\mathbf{\Sigma}$ is given as

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} & \mathbf{\Sigma}_{13} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} & \mathbf{\Sigma}_{23} \\ \mathbf{\Sigma}_{31} & \mathbf{\Sigma}_{32} & \mathbf{\Sigma}_{33} \end{bmatrix}.$$

In a first step, let us focus on the submatrix $\tilde{\mathbf{\Sigma}}_1$ given by

$$\tilde{\mathbf{\Sigma}}_1 = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix} \text{ and partition it as } \tilde{\mathbf{\Sigma}}_1 = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{C} \end{bmatrix},$$

where \mathbf{A} is $r \times r$, \mathbf{C} is $s \times s$ and \mathbf{B} is $r \times s$. Consider the vectors $\mathbf{P} = [P_1, \dots, P_r]'$ and $\mathbf{S} = [S_1, \dots, S_s]'$ defined by

$$\mathbf{P}_t = \sum_{a=0}^{\infty} \alpha_a z_{t-a} e_0, \quad \text{for } t = 1, \dots, r, \quad (4.41)$$

$$\mathbf{S}_t = \sum_{b=0}^{\infty} \delta_b z_{t-b} e_0, \quad \text{for } t = 1, \dots, s. \quad (4.42)$$

It can easily be verified that the covariance matrices of \mathbf{P} and \mathbf{S} , denoted $\mathbf{\Sigma}_{PP}$ and $\mathbf{\Sigma}_{SS}$, are equal to \mathbf{A} and \mathbf{C} . From (4.39), it follows that

$$\begin{aligned} \text{Cov}(P_i, S_j) &= \text{Cov} \left(\sum_{a=0}^{\infty} \alpha_a z_{-i-a} e_0, \sum_{b=0}^{\infty} \delta_b z_{j-b} e_0 \right) \\ &= \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} \alpha_a \delta_b \mathbb{E}(z_{-i-a} z_{j-b} e_0^2) \\ &= \sum_{a=0}^{\infty} \alpha_a \delta_{b+i-j} (\mathcal{J} + \tilde{\mathcal{J}} \sum_{m=1}^q \beta_m^2 \sigma_m^2). \end{aligned}$$

We know that $\mathcal{J} > 1$ by condition (A5) of Andrews et al. (2006). We also have that $\tilde{\mathcal{J}} \sum_{m=1}^q \beta_m^2 \sigma_m^2 = \mathcal{J} \left(\frac{\sum_{m=1}^q \beta_m^2 \sigma_m^2}{\sigma^2} \right) > 0$, which in turn implies that $(\mathcal{J} + \tilde{\mathcal{J}} \sum_{m=1}^q \beta_m^2 \sigma_m^2) > 1$. Similar to Breidt et al. (1991), we exploit that the matrices \mathbf{A} and \mathbf{C} are positive definite since there is no linear dependence within the vectors \mathbf{P} and \mathbf{S} . We proceed by proving the positive definiteness of $\tilde{\Sigma}$ by showing that the Schur Complement of the block \mathbf{A} of the matrix $\tilde{\Sigma}$ given as $\mathbf{C} - \mathbf{B}'\mathbf{A}^{-1}\mathbf{B}$ is positive definite. We know that the covariance matrix of $\mathbf{S} - \Sigma_{SP}\Sigma_{PP}^{-1}$, i.e. $\mathbf{C} - (\mathcal{J} + \tilde{\mathcal{J}} \sum_{m=1}^q \beta_m^2 \sigma_m^2) \mathbf{B}'\mathbf{A}^{-1}\mathbf{B}$, is positive semidefinite and hence for a nonzero vector $\mathbf{c} \in \mathbb{R}^s$ with $\mathbf{B}\mathbf{c} \neq \mathbf{0}$, we have that

$$\mathbf{c}'(\mathbf{C} - \mathbf{B}'\mathbf{A}^{-1}\mathbf{B})\mathbf{c} > \mathbf{c}'(\mathbf{C} - (\mathcal{J} + \tilde{\mathcal{J}} \sum_{m=1}^q \beta_m^2 \sigma_m^2) \mathbf{B}'\mathbf{A}^{-1}\mathbf{B})\mathbf{c} \geq 0.$$

Alternatively, if $\mathbf{B}\mathbf{c} = \mathbf{0}$, we have that

$$\mathbf{c}'(\mathbf{C} - \mathbf{B}'\mathbf{A}^{-1}\mathbf{B})\mathbf{c} = \mathbf{c}'\mathbf{C}\mathbf{c} > 0,$$

by the positive definiteness of \mathbf{C} . Hence, now that we established positive definiteness of $\tilde{\Sigma}_1$, we can repartition the matrix Σ as

$$\Sigma = \begin{bmatrix} \tilde{\Sigma}_1 & \tilde{\Sigma}_2 \\ \tilde{\Sigma}_2' & \tilde{\Sigma}_3 \end{bmatrix},$$

where $\tilde{\Sigma}_1$ is $(r+s) \times (r+s)$, $\tilde{\Sigma}_2 = [\Sigma_{12}, \Sigma_{23}]'$ is $(r+s) \times q$ and $\tilde{\Sigma}_3 = \Sigma_{33}$ is $q \times q$. Since $\Sigma_{33} = \text{diag}(\sigma_1^2, \dots, \sigma_m^2)$, we have that for a nonzero vector $\mathbf{c} \in \mathbb{R}^q$, $\mathbf{c}'\Sigma_{33}\mathbf{c} = c_1^2\sigma_1^2 + \dots + c_m^2\sigma_m^2 > 0$. Hence, as we know that $\tilde{\Sigma}_1$ and $\tilde{\Sigma}_3$ are positive definite, it is sufficient to show that the Schur complement of the block $\tilde{\Sigma}_1$ of the matrix Σ is positive definite, which can be established analogous to the case above. The positive definiteness of Ω follows from condition (A6) in Andrews et al. (2006).

Proof of Theorem 4.1

Similar to Lanne and Saikkonen (2011a), we first present the second partial derivatives of the function $g_t(\boldsymbol{\theta})$. We set $h(x; \boldsymbol{\lambda}) = f'(x; \boldsymbol{\lambda})/f(x; \boldsymbol{\lambda})$, such that

$$h'(x; \boldsymbol{\lambda}) = \frac{f''(x; \boldsymbol{\lambda})}{f(x; \boldsymbol{\lambda})} - \left(\frac{f'(x; \boldsymbol{\lambda})}{f(x; \boldsymbol{\lambda})} \right)^2,$$

which can easily be verified using the quotient rule. Let \mathbf{Y}_t be the $(r \times s)$ matrix with elements y_{t-i+j} . Write $\tilde{v}_t = v_t(\boldsymbol{\varphi})$ and $\tilde{u}_t = u_t(\boldsymbol{\phi})$ and thus $\tilde{\mathbf{V}}_{t-1} = [\tilde{v}_{t-1}, \dots, \tilde{v}_{t-r}]'$ and $\tilde{\mathbf{U}}_{t+1} = [\tilde{u}_{t+1}, \dots, \tilde{u}_{t+s}]'$ to simplify notation. Similarly, $\tilde{\varepsilon}_t = \tilde{v}_t - \phi_1 \tilde{v}_{t-1} - \dots - \phi_r \tilde{v}_{t-r} = \tilde{u}_t - \varphi_1 \tilde{u}_{t+1} - \dots - \varphi_s \tilde{u}_{t+s}$ denotes ε_t evaluated at an arbitrary point in the permissible parameter space, not the true one. Then, the second partial derivatives in the MARX case can be obtained through differentiation, similar to Lanne and Saikkonen (2011a) and Breidt et al. (1991):

$$\begin{aligned}
 \partial^2 g_t(\boldsymbol{\theta}) / \partial \boldsymbol{\phi} \partial \boldsymbol{\phi}' &= \sigma^{-2} h'(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda}) \tilde{\mathbf{V}}_{t-1} \tilde{\mathbf{V}}_{t-1}', \\
 \partial^2 g_t(\boldsymbol{\theta}) / \partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}' &= \sigma^{-2} h'(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda}) \tilde{\mathbf{U}}_{t+1} \tilde{\mathbf{U}}_{t+1}', \\
 \partial^2 g_t(\boldsymbol{\theta}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}' &= \sigma^{-2} h'(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda}) \mathbf{X}_t \mathbf{X}_t', \\
 \partial^2 g_t(\boldsymbol{\theta}) / \partial \sigma^2 &= 2\sigma^{-3} h(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda}) \tilde{\varepsilon}_t + \sigma^{-4} h'(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda}) \tilde{\varepsilon}_t^2 + \sigma^{-2}, \\
 \partial^2 g_t(\boldsymbol{\theta}) / \partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}' &= \frac{1}{f(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda})} \frac{\partial^2 f(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} \\
 &\quad - \frac{1}{f^2(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda})} \left(\frac{\partial f(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} \right) \left(\frac{\partial f(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} \right)', \\
 \partial^2 g_t(\boldsymbol{\theta}) / \partial \boldsymbol{\phi} \partial \boldsymbol{\varphi}' &= \sigma^{-2} h'(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda}) \tilde{\mathbf{V}}_{t-1} \tilde{\mathbf{U}}_{t+1}' + \sigma^{-1} h(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda}) \mathbf{Y}_t, \\
 \partial^2 g_t(\boldsymbol{\theta}) / \partial \boldsymbol{\phi} \partial \boldsymbol{\beta}' &= \sigma^{-2} h'(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda}) \tilde{\mathbf{V}}_{t-1} \mathbf{X}_t', \\
 \partial^2 g_t(\boldsymbol{\theta}) / \partial \boldsymbol{\phi} \partial \sigma &= \sigma^{-3} h'(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda}) \tilde{\varepsilon}_t \tilde{\mathbf{V}}_{t-1} + \sigma^{-2} h(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda}) \tilde{\mathbf{V}}_{t-1}, \\
 \partial^2 g_t(\boldsymbol{\theta}) / \partial \boldsymbol{\phi} \partial \boldsymbol{\lambda}' &= -\sigma^{-1} \tilde{\mathbf{V}}_{t-1} \partial h(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda}) / \partial \boldsymbol{\lambda}', \\
 \partial^2 g_t(\boldsymbol{\theta}) / \partial \boldsymbol{\varphi} \partial \boldsymbol{\beta}' &= \sigma^{-2} h'(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda}) \tilde{\mathbf{U}}_{t+1} \mathbf{X}_t', \\
 \partial^2 g_t(\boldsymbol{\theta}) / \partial \boldsymbol{\varphi} \partial \sigma &= \sigma^{-3} h'(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda}) \tilde{\varepsilon}_t \tilde{\mathbf{U}}_{t+1} + \sigma^{-2} h(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda}) \tilde{\mathbf{U}}_{t+1}, \\
 \partial^2 g_t(\boldsymbol{\theta}) / \partial \boldsymbol{\varphi} \partial \boldsymbol{\lambda}' &= -\sigma^{-1} \tilde{\mathbf{U}}_{t+1} \partial h(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda}) / \partial \boldsymbol{\lambda}', \\
 \partial^2 g_t(\boldsymbol{\theta}) / \partial \boldsymbol{\beta} \partial \sigma &= \sigma^{-3} h'(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda}) \tilde{\varepsilon}_t \mathbf{X}_t + \sigma^{-2} h(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda}) \mathbf{X}_t, \\
 \partial^2 g_t(\boldsymbol{\theta}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\lambda}' &= -\sigma^{-1} \mathbf{X}_t \partial h(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda}) / \partial \boldsymbol{\lambda}', \\
 \partial^2 g_t(\boldsymbol{\theta}) / \partial \sigma \partial \boldsymbol{\lambda}' &= -\sigma^{-2} \tilde{\varepsilon}_t \partial h(\sigma^{-1} \tilde{\varepsilon}_t; \boldsymbol{\lambda}) / \partial \boldsymbol{\lambda}'.
 \end{aligned}$$

It can be verified that $\mathbb{E}(\partial^2 g_t(\boldsymbol{\theta}_0) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}') = -\text{diag}(\boldsymbol{\Sigma}, \boldsymbol{\Omega})$. The proof for consistency is exactly the same as in Lanne and Saikkonen (2011a). That is, similar to Andrews et al. (2006), we use the Taylor expansion

$$\sum_{t=r+1}^{T-s} \left[g_t(\boldsymbol{\theta}_0 + T^{-1/2} \mathbf{c}) - g_t(\boldsymbol{\theta}_0) \right] = \frac{1}{\sqrt{T}} \sum_{t=r+1}^{T-s} \mathbf{c}' \frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \frac{1}{2T} \sum_{t=r+1}^{T-s} \mathbf{c}' \frac{\partial^2 g_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \mathbf{c}$$

$$+ \frac{1}{2T} \sum_{t=r+1}^{T-s} \mathbf{c}' \left(\frac{\partial^2 g_t(\boldsymbol{\theta}_T^*(\mathbf{c}))}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \frac{\partial^2 g_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) \mathbf{c},$$

where $\mathbf{c} \in \mathbb{R}^{p+q+1+d}$ and the argument $\boldsymbol{\theta}_T^*(\mathbf{c})$ in the matrix of second partial derivatives means that each row is evaluated at an intermediate point lying between the true parameter value $\boldsymbol{\theta}_0$ and $T^{-1/2}\mathbf{c}$. If $\|\cdot\|$ denotes the Euclidian norm we have $\sup_{\mathbf{c} \in \mathcal{C}} \|\boldsymbol{\theta}_T^*(\mathbf{c}) - \boldsymbol{\theta}_0\| \rightarrow 0$ for any compact set $\mathcal{C} \subset \mathbb{R}^{p+q+1+d}$. Using the dominance conditions (A7) in Davis et al. (2006), arguments similar to Breidt et al. (1991, p. 186-190) and assumption (A1) in this chapter, it can be shown that a uniform law of large numbers for stationary ergodic processes applies to $\partial^2 g_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'$ over any small enough compact neighborhood $\boldsymbol{\theta}_0$. We can conclude that

$$\frac{1}{T} \sum_{t=r+1}^{T-s} \mathbf{c}' \left(\frac{\partial^2 g_t(\boldsymbol{\theta}_T^*(\mathbf{c}))}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \frac{\partial^2 g_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) \mathbf{c} \xrightarrow{p} 0,$$

for $\mathbf{c} \in \mathcal{C}$. As in the proof of Theorem 1 of Andrews et al. (2006), we can make use of Remark 1 of Davis et al. (1992) and complete the proof.

Chapter 5

Detecting Co-Movements in Noncausal Time Series

This chapter introduces the notion of common noncausal features and proposes tools for detecting the presence of co-movements in economic and financial time series subject to phenomena such as asymmetric cycles and speculative bubbles. For purely causal or noncausal vector autoregressive models with more than one lag, the presence of a reduced rank structure allows to identify causal from non-causal systems using the usual Gaussian likelihood framework. This result cannot be extended to mixed causal-noncausal models, and an approximate maximum likelihood estimator assuming non-Gaussian disturbances is needed for this case. We find common bubbles in both commodity prices and price indicators.¹

¹This chapter is based on the paper Cubadda, G., Hecq, A. and S. Telg (2017), Detecting Co-Movements in Noncausal Time Series, *MPRA Paper 77254*, University Library of Munich, Germany.

5.1 Introduction

This chapter studies the existence of common cyclical features in economic and financial variables by considering both causal and noncausal vector autoregressive (VAR) models. A feature is a dominant characteristic observed in univariate time series. Examples of such features are the presence of a stochastic trend, a dynamic pattern, a conditional time varying volatility process, seasonality, structural breaks, nonlinearities, etc. A feature is said to be common if a linear combination of the series no longer has the feature even though each of the series individually has it. One will then talk about cointegration, common cyclical features, common ARCH, co-breaking, common seasonality, and co-jumps to name a few. In this article, we focus on common cyclical features as introduced by Engle and Kozicki (1993) and Vahid and Engle (1993). That is, we are interested in the existence of commonalities in the dynamics of time series. Those restrictions have implications for forecasting accuracy, parameter efficiency as well as for the interpretation of business cycle co-movements and the evaluation of economic theories (Guillén et al., 2015; Issler and Vahid, 2001).

Although co-movements in the cyclical fluctuations of many economic time series seem theoretically well motivated and even graphically visible, it often happens that the presence of common dynamics is rejected by formal test statistics. The first obvious reason stems from the fact that cycles do not have to be synchronous and therefore several statistical models have been introduced to account for adjustment delays.² Several additional reasons have been suggested in the literature, among which the impact of misspecifications: presence of conditional heteroskedasticity (Candelon et al., 2005), seasonal adjustment (Hecq, 1998; Cubadda, 1999), outlying observations and nonlinear phenomena.

In this chapter we propose a novel explanation for not being able to detect common cyclical patterns as often as expected. In fact, one of the most popular approaches to test for serial correlation common features, SCCF henceforth, is to consider the VAR model of order p of a stationary n -vector time series \mathbf{Y}_t , denoted as VAR(p)

$$\Pi(L)\mathbf{Y}_t \equiv (\mathbf{I}_n - \Pi_1 L - \dots - \Pi_p L^p)\mathbf{Y}_t = \boldsymbol{\varepsilon}_t,$$

²Those extensions include the polynomial serial correlation common feature specification (Cubadda and Hecq, 2001), the weak form common feature model (Hecq et al., 2006) and the codependent cycle approach (Vahid and Engle, 1997). See Cubadda (2007) for a unifying framework of the various forms of common features.

where L is the lag operator, and ε_t are *iid* innovations with $\mathbb{E}(\varepsilon_t) = \mathbf{0}$ and $\mathbb{E}(\varepsilon_t \varepsilon_t') = \mathbf{\Omega}$ (positive definite). Under the existence of SCCF, the $\text{VAR}(p)$ can be written as

$$(\mathbf{I}_n - \delta_{\perp} \beta_1' L - \dots - \delta_{\perp} \beta_p' L^p) \mathbf{Y}_t = \varepsilon_t,$$

where δ_{\perp} is a full rank $n \times (n - k)$ matrix ($0 < k < n$) such that there exist k linear combinations $\delta' \mathbf{Y}_t = \delta' \varepsilon_t$ that annihilate the entire dynamics (where δ' is such that $\delta' \delta_{\perp} = \mathbf{0}$). Hence, a reduced rank regression approach on the matrix $[\mathbf{\Pi}_1, \dots, \mathbf{\Pi}_p]$ is typically used to detect the commonalities among series \mathbf{Y}_t . However, the linear VAR model cannot capture several important characteristics of macroeconomic and financial variables, such as asymmetric cycles or bubbles phenomena. Noncausal VARs, i.e. VAR models with leads of series \mathbf{Y}_t in place of their lags or, more generally, mixed causal-noncausal VARs, which are models where both lags and leads of \mathbf{Y}_t are present in the right-hand side (see Lanne and Saikkonen, 2013), are able to generate much richer dynamics than usual VARs.³

Causal and noncausal VARs are often hardly visually distinguishable (e.g., see the graphs in Section 5.4) and non-Gaussian distributions are needed to discriminate between those models by estimation techniques like Maximum Likelihood (ML). We show in this chapter that for purely causal or noncausal VAR models with more than one lag or lead, the presence of a reduced rank structure in the VAR coefficient matrices allows to identify causal from noncausal systems even using the Gaussian framework. Using either lags or leads within a canonical correlation analysis or a Generalized Method of Moments (GMM) approach, the reduced rank restrictions help to identify the correct model. This result stems from the fact that, except for VARs with either one lag or one lead only, the existence of SCCF implies that the autocorrelation matrices of series \mathbf{Y}_t have a common left null space for either any lag or any lead different from zero. However, this important observation does not extend to mixed causal-noncausal models.

The rest of the chapter is organized as follows. Section 5.2 summarizes results about the mixed causal-noncausal autoregressive model with common features. Section 5.3 introduces our testing strategy for identifying purely causal and noncausal models with SCCF. The finite sample behavior of reduced rank regressions and GMM orthogonality tests when noncausality is present is studied in Section

³More precisely, we define a noncausal model as a model which has a unique stationary solution in terms of current and future error terms. For the mixed causal-noncausal model, this is the two-sided MA representation, i.e., past, current and future disturbances.

5.4. Section 5.5 applies our new modelling procedure to investigate the presence of common bubbles in price index variables as well as in commodity prices. Section 5.6 summarizes and concludes.

5.2 Causal and Noncausal Models

Mixed causal-noncausal models have recently become increasingly popular because of their appealing properties. On a theoretical level, they offer the possibility to rewrite a process with explosive roots in direct time into as process in reverse time with roots outside the unit circle. In the applied econometric literature, there is a growing interest because mixed causal-noncausal models (*i*) might improve forecast performances, (*ii*) can be solutions to rational expectations models which take forward-looking behavior into account and (*iii*) are able to model nonlinear features in macroeconomic and financial data. A more extensive overview and the appropriate references can be found in Hecq et al. (2016), Lanne et al. (2012a), Lanne and Saikkonen (2011a, 2011b), Hencic and Gouriéroux (2014), Gouriéroux and Zakoïan (2016).

5.2.1 Univariate Models

The mixed causal-noncausal autoregressive process for a scalar series y_t , denoted as MAR(r, s), can be represented as

$$\pi(L)\phi(L^{-1})y_t \equiv (1 - \pi_1 L - \dots - \pi_r L^r)(1 - \phi_1 L^{-1} - \dots - \phi_s L^{-s})y_t = \epsilon_t, \quad (5.1)$$

where ϵ_t are scalar *iid* innovations. Notice that the backshift operator L creates lags when raised to positive powers and leads when raised to negative powers, i.e., $L^j y_t = y_{t-j}$ and $L^{-j} y_t = y_{t+j}$. When $\phi_1 = \dots = \phi_s = 0$, the process y_t is a purely causal MAR($r, 0$), also known as the conventional causal AR(r) process

$$\pi(L)y_t = \epsilon_t,$$

while the process is a purely noncausal MAR($0, s$) model

$$\phi(L^{-1})y_t = \epsilon_t,$$

when $\pi_1 = \dots = \pi_r = 0$.

In order to ensure stationarity, the roots of both polynomials $\pi(L)$ and $\phi(L^{-1})$ are assumed to lie outside the unit circle. Additionally, the identifiability of the causal and the noncausal part is established by assuming that the error terms ϵ_t are non-Gaussian with $\mathbb{E}(|\epsilon_t|^\delta) < \infty$ for $\delta > 0$ (Breidt et al., 1991). Most articles by Lanne, Saikkonen and coauthors use a Student's t distribution as an alternative for the Gaussian distribution while Gouriéroux and coauthors rely on α -stable distributions (in practice, mostly Cauchy or mixtures of Cauchy and Gaussian).

It is well known that one can define a pseudo-causal AR(p) representation of (5.1) with $p = r + s$ lags⁴

$$a(L)y_t = \epsilon_t^*, \tag{5.2}$$

where ϵ_t^* is a white noise error, and the autoregressive polynomial $a(L)$ has all roots outside the unit circle (see, Brockwell and Davis, 1991; Hecq et al., 2016). Although not autocorrelated, the error term ϵ_t^* is generally not *iid* anymore. Gouriéroux and Zakoïan (2016) exploit this property to discriminate between purely causal and noncausal AR(1) processes.

5.2.2 Noncausality and Common Features

The main idea of this chapter can be characterized well by an example. Let us consider the monthly price levels of a barrel of oil for Brent and West Texas Intermediate (WTI) from May 1987 until September 2017. These series, downloaded from the U.S. Energy Information Administration, are displayed in Figure 5.1. Graphically, the two series appear extremely similar with only a few differences. The most notable one is that Brent is systematically higher than WTI during the period 2011-2014. In general, however, the series show strong co-movements with several episodes displaying bubble features.

Gouriéroux and Zakoïan (2016) show that the noncausal autoregressive model with a fat-tailed error distribution is able to capture such dynamics, whereas its causal counterpart cannot. Hence, before we focus on detecting common features, let us consider the univariate modelling of the series in Figure 5.1. The proper detrending of noncausal processes is troublesome, as the series appear locally explosive but globally stationary. For this reason, it is difficult to test whether such series are trend- or difference-stationary. Saikkonen and Sandberg (2016)

⁴Lanne and Saikkonen (2011a) propose to use usual information criteria to first get the pseudo-causal AR(p). Hencic and Gouriéroux (2014) propose to look at the empirical ACF which is consistent even in the infinite variance case to determine p .

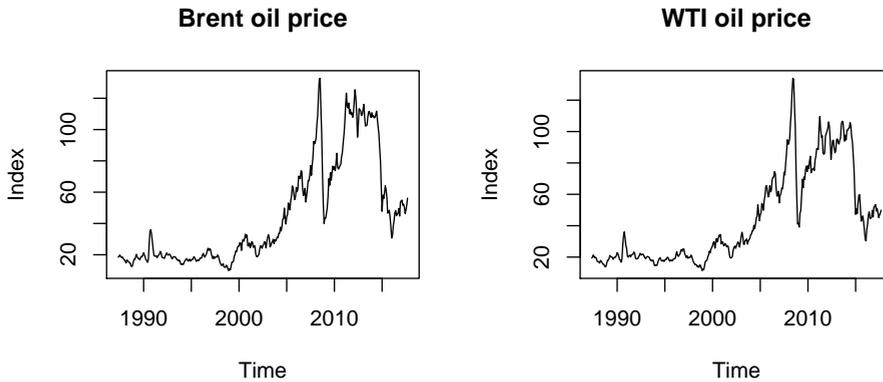


Figure 5.1: Brent and WTI crude oil prices

show that the ADF test suffers from power and size distortions in this framework. Several papers (e.g., Hencic and Gouriéroux, 2014; Cavaliere et al., 2017) apply polynomial detrending on the series of interest prior to fitting noncausal models. However, there is no clear reason to believe that one method is better than the other. We do not want to address this issue here and simply apply standard analysis to illustrate the main point. The ADF test on the price levels of Brent and WTI fails to reject the null of a unit root. As a result, we perform model selection on the growth rates of the prices. For both series, the pseudo-causal model selected is the AR(1). Using the Student's t approximate maximum likelihood (AML) estimator proposed by Lanne and Saikkonen (2011a), we identify a noncausal AR(1) for Brent and a causal AR(1) for WTI. However, as the estimated degrees of freedom are close to 10 for WTI, identification is weak in finite samples as shown in Hecq et al. (2016). This is also reflected in the values of the log-likelihood of the causal AR(1) and noncausal AR(1) for WTI, which are extremely close. This is probably due to the fact that a major part of the bubble pattern present in the levels disappears when considering the growth rates.

However, if we now continue to model both processes jointly as a bivariate VAR(p), information criteria (AIC, BIC, HQ) indicate a dynamic order p equal to 2 or 4. The tests for common features proposed in the sequel of this chapter are performed on the causal as well as the noncausal VAR(2) and VAR(4). In both

cases, common features are only detected in the noncausal model.⁵ The intuitive explanation for this is that the dynamics of both series are indeed noncausal of nature. For this reason, the causal VAR does not detect these commonalities, while the noncausal VAR does. In the following sections, we discuss the notion of common features in the multivariate framework, allowing for noncausality, more formally.

5.2.3 Multivariate Models with Common Features

Consider a mixed vector autoregressive model of order (r, s) , denoted as $MVAR(r, s)$, of a second-order stationary n -vector time series \mathbf{Y}_t such that

$$\Pi(L)\Phi(L^{-1})\mathbf{Y}_t \equiv (\mathbf{I}_n - \Pi_1 L - \dots - \Pi_r L^r)(\mathbf{I}_n - \Phi_1 L^{-1} - \dots - \Phi_s L^{-s})\mathbf{Y}_t = \boldsymbol{\varepsilon}_t, \quad (5.3)$$

where $\Pi(L)$ and $\Phi(L^{-1})$ are $n \times n$ polynomial matrices with both $\det\{\Pi(L)\} = 0$ and $\det\{\Phi(L^{-1})\} = 0$ having solutions outside the unit circle. The error term $\boldsymbol{\varepsilon}_t$ is an *iid* innovation process with $\mathbb{E}(\boldsymbol{\varepsilon}_t) = \mathbf{0}$ and $\mathbb{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \boldsymbol{\Omega}$ (positive definite). From (5.3) it follows that the series \mathbf{Y}_t have a strictly stationary solution as a two-sided MA representation

$$\mathbf{Y}_t = \Psi(L, L^{-1})\boldsymbol{\varepsilon}_t = \sum_{j=-\infty}^{\infty} \Psi_j \boldsymbol{\varepsilon}_{t-j}, \quad (5.4)$$

where $\Psi(L, L^{-1}) \equiv [\Pi(L)\Phi(L^{-1})]^{-1}$, with $\Psi_0 = \mathbf{I}_n$ and Ψ_j converging to zero at a geometric rate as $j \rightarrow \pm\infty$.

To extend the SCCF property to the $MVAR(r, s)$ case, we search for the existence of a full rank matrix $\boldsymbol{\delta}$ such that $\mathbb{E}(\boldsymbol{\delta}'\mathbf{Y}_t | \mathbf{Y}_{t+s}, \dots, \mathbf{Y}_{t+1}, \mathbf{Y}_{t-1}, \dots, \mathbf{Y}_{t-r}) = \mathbf{0}$. This gives rise to the following definition.

Definition 5.1. Series \mathbf{Y}_t have k common features, CFs henceforth, if there exists an $n \times k$ ($0 < k < n$) full rank matrix $\boldsymbol{\delta}$ such that $\boldsymbol{\delta}'\Pi(L)\Phi(L^{-1}) = \boldsymbol{\delta}'$ or, equivalently, $\boldsymbol{\delta}'\mathbf{Y}_t = \boldsymbol{\delta}'\boldsymbol{\varepsilon}_t$. We also have that $\boldsymbol{\delta}'\tilde{\Phi}(L^{-1})\tilde{\Pi}(L) = \boldsymbol{\delta}'$ although the product of matrices is not commutative.⁶

⁵Performing the same analysis on polynomial detrended series does not change these results qualitatively.

⁶Notice that, differently from the univariate case, representation (5.3) is not unique. Since the polynomial matrices are generally not commutative, one may alternatively use the model $\tilde{\Phi}(L^{-1})\tilde{\Pi}(L)\mathbf{Y}_t \equiv (\mathbf{I}_n - \tilde{\Phi}_1 L^{-1} - \dots - \tilde{\Phi}_s L^{-s})(\mathbf{I}_n - \tilde{\Pi}_1 L - \dots - \tilde{\Pi}_r L^r)\mathbf{Y}_t = \boldsymbol{\varepsilon}_t$. Gouriéroux and Jasiak (2017b) have a different representation for the MVAR which does not involve the multiplicative structure of two coefficient matrices.

Remark 5.1. By premultiplying both sides of Equation (5.3) by δ' and in view of the definition of CFs we have that

$$\delta' \Psi(L, L^{-1}) = \delta',$$

which tells us that the bilateral linear impulse response function of series \mathbf{Y}_t is collinear at any lag/lead different from zero. It is easy to check that the reverse implication applies as well.

5.2.4 Co-Movements in Purely Causal or Noncausal Models

We first focus on the two polar cases in which data are generated by either a purely causal $\text{MVAR}(r, 0)$ or a purely noncausal $\text{MVAR}(0, s)$ model. Then we illustrate the properties of the usual common cyclical feature test statistics, e.g., canonical correlations, for noncausal models. Finally, we show that the mixed model necessitates to use a non-Gaussian likelihood framework even when a reduced rank structure is present.

Definition 5.2. In purely causal [noncausal] models with $\Phi(L^{-1}) = \mathbf{I}_n$ [$\Pi(L) = \mathbf{I}_n$], series \mathbf{Y}_t have k causal [noncausal] common features, CCFs [NCCFs] henceforth, if there exists an $n \times k$ ($0 < k < n$) full rank matrix δ_c [δ_{nc}] such that $\delta'_c \Pi(L) = \mathbf{0}$ [$\delta'_{nc} \Phi(L^{-1}) = \mathbf{0}$]. The common cyclical feature model in the purely causal [noncausal] case implies $\mathbb{E}(\delta'_c \mathbf{Y}_t | \mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \dots) = \mathbf{0}$ [$\mathbb{E}(\delta'_{nc} \mathbf{Y}_t | \mathbf{Y}_{t+1}, \mathbf{Y}_{t+2}, \dots) = \mathbf{0}$]. The $n \times (n - k)$ orthogonal complement of matrix δ_c [δ_{nc}] is denoted as $\delta_{c\perp}$ [$\delta_{nc\perp}$].

Remark 5.2. The usual SCCF in a $\text{VAR}(r)$ is obviously equivalent to the CCF in a $\text{MVAR}(r, 0)$.

We start the analysis by concentrating on the cases in which data are generated by a purely noncausal $\text{MVAR}(0, s)$ model. Assuming that the matrix δ_{nc} exists (we skip the indexes nc and c to simplify notations when no confusion is possible). The noncausal VAR model then reads as

$$\mathbf{Y}_t = \sum_{j=1}^s \Phi_j \mathbf{Y}_{t+j} + \varepsilon_t = \delta_{\perp} \sum_{j=1}^s \mathbf{A}'_j \mathbf{Y}_{t+j} + \varepsilon_t, \quad (5.5)$$

where $\delta'_{\perp} \delta = \mathbf{0}$ and at least one of the $n \times (n - k)$ matrices \mathbf{A}_j , for $j = 1, \dots, s$, has full rank. As shown by Lanne and Saikkonen (2013), series \mathbf{Y}_t admit a causal autoregressive representation $\text{MVAR}(s, 0)$ with the same spectral density as (5.5)

and with errors that are multivariate white noise but not innovations with respect to the past. When no CFs exist, this implies that the noncausal and the pseudo-causal representation cannot be distinguished by means of the Gaussian likelihood framework.

We wish to examine the implications of the presence of such CFs in the pseudo-causal VAR representation of the system. In order to do that, we start by defining autocovariances $\mathbf{\Gamma}_y(-j) = \mathbb{E}(\mathbf{Y}_t \mathbf{Y}'_{t+j})$ and note that

$$\delta' \mathbf{\Gamma}_y(-j) = \delta' \mathbb{E}(\mathbf{Y}_t \mathbf{Y}'_{t+j}) = \delta' \mathbb{E}(\boldsymbol{\varepsilon}_t \mathbf{Y}'_{t+j}) = \mathbf{0}, \quad \forall j > 0,$$

as $\delta' \mathbf{Y}_t = \delta' \boldsymbol{\varepsilon}_t$. Hence, it follows that we can factorize these autocovariances as the product

$$\mathbf{\Gamma}_y(-j) = \boldsymbol{\delta}_\perp \boldsymbol{\Upsilon}'_y(-j), \quad \forall j > 0,$$

where $\boldsymbol{\Upsilon}_y(-j)$ is $(n-k) \times n$ matrix. Let us now rewrite the noncausal VAR in (5.5) under reduced rank restrictions as

$$\mathbf{Y}_t = \boldsymbol{\Phi}' \mathbf{X}_t + \boldsymbol{\varepsilon}_t = \boldsymbol{\delta}_\perp \mathbf{A}' \mathbf{X}_t + \boldsymbol{\varepsilon}_t, \quad (5.6)$$

where $\mathbf{X}'_t = [\mathbf{Y}'_{t+1}, \dots, \mathbf{Y}'_{t+s}]$, $\boldsymbol{\Phi}' = [\boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_s]$ and $\mathbf{A}' = [\mathbf{A}'_1, \dots, \mathbf{A}'_s]$. The coefficient matrix $\boldsymbol{\Phi}$ is linked to the autocorrelation matrix function $\mathbf{\Gamma}_y(-j)$ through the relation

$$\begin{aligned} \boldsymbol{\Phi}' &= \mathbb{E}(\mathbf{Y}_t \mathbf{X}'_t) [\mathbb{E}(\mathbf{X}_t \mathbf{X}'_t)]^{-1} = [\mathbf{\Gamma}_y(-1), \dots, \mathbf{\Gamma}_y(-s)] [\boldsymbol{\Gamma}_x(0)]^{-1} \\ &= \boldsymbol{\delta}_\perp [\boldsymbol{\Upsilon}'_y(-1), \dots, \boldsymbol{\Upsilon}'_y(-s)] [\boldsymbol{\Gamma}_x(0)]^{-1}, \end{aligned} \quad (5.7)$$

where

$$\boldsymbol{\Gamma}_x(0) = \begin{bmatrix} \mathbf{\Gamma}_y(0) & \mathbf{\Gamma}_y(-1) & \cdots & \mathbf{\Gamma}_y(-s+1) \\ \mathbf{\Gamma}_y(1) & \mathbf{\Gamma}_y(0) & \cdots & \mathbf{\Gamma}_y(-s+2) \\ \vdots & & & \vdots \\ \mathbf{\Gamma}_y(s-1) & \mathbf{\Gamma}_y(s-2) & \cdots & \mathbf{\Gamma}_y(0) \end{bmatrix},$$

and hence $\delta' \Phi' = \mathbf{0}$. Let us now write the pseudo-causal representation of series \mathbf{Y}_t as follows

$$\mathbf{Y}_t = \sum_{j=1}^s \tilde{\Phi}_j \mathbf{Y}_{t-j} + \tilde{\varepsilon}_t = \tilde{\Phi}' \mathbf{Z}_t + \tilde{\varepsilon}_t, \quad (5.8)$$

where $\mathbf{Z}'_t = [\mathbf{Y}'_{t-1}, \dots, \mathbf{Y}'_{t-s}]$ and $\tilde{\varepsilon}_t$ is a multivariate white noise process such that $\mathbb{E}(\tilde{\varepsilon}_t \mathbf{Y}'_{t-j}) = \mathbf{0}$ for any $j > 0$. The coefficient matrix $\tilde{\Phi}' = [\tilde{\Phi}_1, \dots, \tilde{\Phi}_s]$ of the pseudo-causal VAR is linked to the autocorrelation matrix function of series \mathbf{Y}_t through the relation

$$\begin{aligned} \tilde{\Phi}' &= \mathbb{E}(\mathbf{Y}_t \mathbf{Z}'_t) [\mathbb{E}(\mathbf{Z}_t \mathbf{Z}'_t)]^{-1} \\ &= [\boldsymbol{\Gamma}_y(1), \dots, \boldsymbol{\Gamma}_y(s)] [\boldsymbol{\Gamma}_z(0)]^{-1} = [\boldsymbol{\Upsilon}_y(1) \boldsymbol{\delta}'_{\perp}, \dots, \boldsymbol{\Upsilon}_y(s) \boldsymbol{\delta}'_{\perp}] [\boldsymbol{\Gamma}_z(0)]^{-1}, \end{aligned} \quad (5.9)$$

where

$$\boldsymbol{\Gamma}_z(0) = \begin{bmatrix} \boldsymbol{\Gamma}_y(0) & \boldsymbol{\Gamma}_y(1) & \cdots & \boldsymbol{\Gamma}_y(s-1) \\ \boldsymbol{\Gamma}_y(-1) & \boldsymbol{\Gamma}_y(0) & \cdots & \boldsymbol{\Gamma}_y(s-2) \\ \vdots & & & \vdots \\ \boldsymbol{\Gamma}_y(-s+1) & \boldsymbol{\Gamma}_y(-s+2) & \cdots & \boldsymbol{\Gamma}_y(0) \end{bmatrix},$$

which implies that $\delta' \tilde{\Phi}' \neq \mathbf{0}$. In view of (5.9), we conclude that the pseudo-causal VAR in (5.8) generally does not have any (pseudo) CFs. It is easy to see that the reverse is also true, i.e., the noncausal representation of a causal VAR with CFs does not generally exhibit the (pseudo) CFs.

A special case occurs when $p = 1$. Equations (5.7) and (5.9) simplify to

$$\begin{aligned} \Phi' &= \boldsymbol{\Gamma}_y(-1) [\boldsymbol{\Gamma}_y(0)]^{-1} = \boldsymbol{\delta}_{\perp} \boldsymbol{\Upsilon}'_y(-1) [\boldsymbol{\Gamma}_x(0)]^{-1}, \\ \tilde{\Phi}' &= \boldsymbol{\Gamma}_y(1) [\boldsymbol{\Gamma}_y(0)]^{-1} = \boldsymbol{\Upsilon}_y(1) \boldsymbol{\delta}'_{\perp} [\boldsymbol{\Gamma}_z(0)]^{-1}, \end{aligned}$$

from which we see that the existence of k CFs in a MVAR(0, 1) implies the presence of k CFs in the pseudo-causal representation of series \mathbf{Y}_t . It is easy to check that the reverse is true as well. We summarize the above results in the following proposition.

Proposition 5.1. If series \mathbf{Y}_t are generated by a second-order stationary MVAR($r, 0$) [MVAR(0, s)] with $r > 1$ [$s > 1$], the existence of k CFs does not imply the pres-

ence of any CF in the pseudo-noncausal [pseudo-causal] representation of series \mathbf{Y}_t .

With regard to statistical inference, a relevant implication of the above proposition is that, except for the case of VAR models with either one lead or lag, the Gaussian ML approach is able to identify the unique representation of series \mathbf{Y}_t that exhibit the common features. As is well known, the Gaussian likelihood analysis of the reduced rank regression model (5.6) is based on canonical correlations between \mathbf{Y}_t and \mathbf{X}_t , denoted as $\text{CanCor}(\mathbf{Y}_t, \mathbf{X}_t)$; see e.g. Johansen (2008) and the references therein.

Remark 5.3. In case of the MVAR(1, 0) and MVAR(0, 1) model, it is impossible to discriminate between causal and noncausal CFs using the Gaussian ML analysis. Indeed, by the assumption of stationarity and noting that autocorrelation is a symmetric measure, it follows that

$$\text{CanCor}(\mathbf{Y}_t, \mathbf{Y}_{t+1}) = \text{CanCor}(\mathbf{Y}_{t-1}, \mathbf{Y}_t) = \text{CanCor}(\mathbf{Y}_t, \mathbf{Y}_{t-1}).$$

Consequently, even if there exists a NCCF, we will also detect a CCF using canonical correlations. However, although the eigenvalues from the solutions of $\text{CanCor}(\mathbf{Y}_t, \mathbf{Y}_{t+1})$ and $\text{CanCor}(\mathbf{Y}_t, \mathbf{Y}_{t-1})$ are identical, the eigenvectors, and hence the relationships that link series, will be different. This might explain why “strange” estimates of common cycle relationships are sometimes obtained in empirical works.

5.2.5 Co-Movements in Mixed Models

Let us now consider the general case of a MVAR(r, s) where both r and s are larger than zero. As shown by Lanne and Saikkonen (2013), series \mathbf{Y}_t admit a pseudo-causal MVAR($p, 0$) representation, where $p = r + s$, with the same spectral density as (5.3) and with errors that are multivariate white noise but not innovations with respect to the past. In view of (5.4), we see that when both r and s are positive, the autocorrelation matrix function of series \mathbf{Y}_t reads

$$\Gamma_y(j) = \sum_{i=-\infty}^{\infty} \Psi_{i+j} \Omega \Psi'_i = \Omega \Psi'_{-j} + \sum_{\forall i \neq -j} \Psi_i \Omega \Psi'_{i-j},$$

which implies that the coefficient matrix of the corresponding pseudo-causal VAR

$$[\tilde{\Phi}_1, \dots, \tilde{\Phi}_p] = [\Gamma_y(1), \dots, \Gamma_y(p)][\Gamma_z(0)]^{-1},$$

will generally be full rank. It is easy to see that the same conclusion applies to the pseudo-noncausal representation as well. We summarize the above results in the following proposition.

Proposition 5.2. If series \mathbf{Y}_t are generated by a second-order stationary MVAR(r, s) with $r > 0$ and $s > 0$, the existence of k CFs does not imply the presence of any CFs in the pseudo-noncausal [pseudo-causal] representation of series \mathbf{Y}_t .

Note that a simple solution to detecting CFs in mixed models such as the canonical correlation $\text{CanCor}(\mathbf{Y}_t, \mathbf{U}_t)$, where $\mathbf{U}_t' = [\mathbf{X}_t', \mathbf{Z}_t']$, cannot be adopted. The reason is that in the associated reduced rank regression model

$$\mathbf{Y}_t = \delta_{\perp} \mathbf{C}' \mathbf{U}_t + \Theta_0^{-1} \epsilon_t,$$

where \mathbf{C} is a full rank $np \times (n - k)$ matrix and $\Theta_0 \equiv \sum_{j=-\min(r,s)}^{\min(r,s)} \Pi_j \Phi_j$, the regressors are correlated with the errors.

5.3 Test Statistics

In order to test for common cyclical features in purely causal or noncausal models, we first determine the lag order p in standard VAR models by means of various information criteria, see Section 5.4.1 for details. Even in the presence of a noncausal component, we obtain an estimate of the pseudo-causal order $p = s$. Having fixed p , we perform both Likelihood Ratio (LR) and GMM tests to detect the presence of co-movements both in purely causal and noncausal models.

5.3.1 LR Tests

The Gaussian LR test is based on the partial canonical correlations between \mathbf{Y}_t and \mathbf{W}_t , where

$$\mathbf{W}_t = \begin{cases} [\mathbf{Y}'_{t-1}, \dots, \mathbf{Y}'_{t-p}]' & \text{for the causal VAR,} \\ [\mathbf{Y}'_{t+1}, \dots, \mathbf{Y}'_{t+p}]' & \text{for the noncausal VAR,} \end{cases}$$

having concentrated out the effects of the deterministic terms \mathbf{D}_t (e.g., intercepts and dummies). This procedure is denoted as

$$\text{CanCor} \{ \mathbf{Y}_t, \mathbf{W}_t \mid \mathbf{D}_t \}. \quad (5.10)$$

The LR test for null hypothesis that there exists at least k common feature vectors is based on the statistic

$$\xi_{LR} = -T \sum_{i=1}^k \ln(1 - \hat{\lambda}_i), \quad k = 1, 2, \dots, n,$$

where $\hat{\lambda}_i$ is the i -th smallest squared canonical correlations computed from (5.10). More specifically, $\hat{\lambda}_i$ is the i -th smallest eigenvalue of

$$\hat{\Sigma}_{WW}^{-1/2} \hat{\Sigma}_{WY} \hat{\Sigma}_{YY}^{-1} \hat{\Sigma}_{YW} \hat{\Sigma}_{WW}^{-1/2},$$

where $\hat{\Sigma}_{AB}$ denotes the sample covariance matrix of two vector time series \mathbf{A}_t and \mathbf{B}_t . Under the null hypothesis, ξ_{LR} follows asymptotically a $\chi^2(\nu)$ distribution with $\nu = knp - k(n - k)$.

5.3.2 GMM Tests

When the time series \mathbf{Y}_t is bivariate to simplify notation, i.e., $\mathbf{Y}'_t = [y_{1t}, y_{2t}]$, a GMM single equation common feature test can be conducted. First, we write $\boldsymbol{\delta} = [1, -\alpha]'$ and estimate α by the instrumental variable (IV) estimator using \mathbf{W}_t as instruments. The IV estimator of α is given by

$$\hat{\alpha}_{IV} = (\mathbf{y}'_2 \mathbf{W} (\mathbf{W}' \mathbf{W})^{-1} \mathbf{W}' \mathbf{y}_2)^{-1} (\mathbf{X}' \mathbf{W} (\mathbf{W}' \mathbf{W})^{-1} \mathbf{W}' \mathbf{y}_1),$$

where $[\mathbf{y}_1, \mathbf{y}_2] \equiv \mathbf{Y}$ and \mathbf{W} respectively indicate the matrices of $T - p$ observations of variables \mathbf{Y}_t and \mathbf{W}_t after having removed the linear influence of the deterministic terms \mathbf{D}_t . Second, we run an overidentification J -test (Hansen, 1982) for the validity of the orthogonality conditions, i.e. the existence of a common feature relationship, using the statistic

$$J_1(\hat{\alpha}_{IV}) = (\mathbf{u}' \mathbf{W}) (\hat{\sigma}_u^2 \mathbf{W}' \mathbf{W})^{-1} (\mathbf{W}' \mathbf{u}),$$

where $\mathbf{u} = \mathbf{y}_1 - \mathbf{y}_2 \hat{\alpha}_{IV}$ and $\hat{\sigma}_u^2 = \mathbf{u}' \mathbf{u} / (T - p)$.

The tests presented so far embed the assumption of homoscedasticity. Candelon et al. (2005) have illustrated in a Monte Carlo exercise that $J_1(\hat{\alpha}_{IV})$ has large size distortions in the presence of GARCH disturbances. Therefore, we also use the White heteroskedasticity robust estimator (see, e.g., Hamilton, 1994):

$$\hat{\alpha}_{IV-W} = (\mathbf{y}'_2 \mathbf{W}(\mathbf{W}'\mathbf{B}\mathbf{W})^{-1} \mathbf{W}'\mathbf{y}_2)^{-1} (\mathbf{y}'_2 \mathbf{W}(\mathbf{W}'\mathbf{B}\mathbf{W})^{-1} \mathbf{W}'\mathbf{y}_1), \quad (5.11)$$

where $\mathbf{B} = \text{diag}(u_1^2, u_2^2, \dots, u_T^2)$ and $u_t = y_{1t} - \hat{\alpha}'_{IV} y_{2t}$. Then we run a J -test robust to heteroskedasticity using the statistic

$$J_2(\hat{\alpha}_{IV-W}) = (\mathbf{u}'^* \mathbf{W})(\mathbf{W}'\mathbf{B}\mathbf{W})^{-1} (\mathbf{W}'\mathbf{u}^*). \quad (5.12)$$

where $\mathbf{u}^* = \mathbf{y}_1 - \mathbf{y}_2 \hat{\alpha}_{IV-W}$.⁷ The common feature null hypothesis consists of an orthogonality condition between a combination of variables and the instruments contained in \mathbf{W}_t . Under this null hypothesis, both test statistics $J_1(\hat{\alpha}_{IV})$ and $J_2(\hat{\alpha}_{IV-W})$ follow asymptotically a $\chi^2(2p-1)$ in the bivariate case with k , the number of common features, equal to 1.

5.4 Monte Carlo Results

5.4.1 The Pseudo-Causal VAR(p) Order

We first investigate by Monte Carlo simulations the frequency with which various information criteria select the correct order p of the pseudo-causal VAR(p). The data generating process is a MVAR(1,2) $(1 - \mathbf{\Pi}_1 L)(1 - \mathbf{\Phi}_1 L^{-1} - \mathbf{\Phi}_2 L^{-2})\mathbf{Y}_t = \boldsymbol{\varepsilon}_t$, with the following coefficient matrices

$$\mathbf{\Pi}_1 = \begin{bmatrix} 0.4 & 0 \\ 0 & 0.7 \end{bmatrix}, \quad (5.13)$$

$$\mathbf{\Phi}_1 = \begin{bmatrix} 0.5 & 0.2 \\ 0.5 & 0.2 \end{bmatrix}, \quad \mathbf{\Phi}_2 = \begin{bmatrix} -0.35 & 0.25 \\ -0.35 & 0.25 \end{bmatrix}. \quad (5.14)$$

Note that the presence of reduced rank matrices in the noncausal part (5.14) has no implications for the order of the pseudo-causal VAR representation. We

⁷Alternative approaches for handling heteroskedastic disturbances have been evaluated in Hecq and Issler (2012).

consider two relatively small sample sizes, $T = \{100, 250\}$.⁸ For ε_t , we consider the following correlated distributions in addition to the multivariate $\mathcal{N}(0, 1)$: (i) Student's $t(\nu) = \zeta_t \odot (\nu/\chi^2(\nu))^{1/2}$ and (ii) Laplace = $\zeta_t \odot (-\ln(U[0, 1]))^{1/2}$, where \odot denotes the Hadamard product. The ζ_t are simulated according to

$$\zeta_t \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 10 & 3 \\ 3 & 2 \end{bmatrix} \right). \quad (5.15)$$

For (i) we consider the following values for the degrees of freedom $\nu = \{10, 3, 1\}$, where $\nu = 1$ corresponds to the Standard Cauchy case. Each distribution has its own seed. The four information criteria are AIC, BIC, HQ and the Takeuchi information criterion (TIC).⁹ We let them find their minimum for $p = 0$ to $p_{\max} = 8$.

Table 5.1 reports the results over 10,000 replications. Overall, information criteria determine the pseudo-causal VAR order rather well for $T = 250$ (in particular HQ). As usual BIC detects a proportion of models with too parsimonious dynamics. This is clearly a problem for the identification of mixed causal-noncausal models because r lags and s leads are going to be estimated within that p . AIC tends to overestimate the number of lags. The TIC does not perform very well by construction when the kurtosis does not exist, e.g., the $t(3)$. The frequency with which the correct model is found is approaching 100% when $T > 250$ for instance (except for the Cauchy).

5.4.2 Testing for Common Cyclical Features

The previous subsection has illustrated that usual information criteria can guide us towards the right choice for the pseudo-causal VAR order p . Let us now fo-

⁸We simulate an additional 50×2 observations (both sides) burn in period to initialize processes.

⁹The TIC has been proposed by Takeuchi (1976) as an alternative to the AIC as it does not rely on the assumption that the true model is included in the set of investigated specifications. He obtained the following criterion

$$\text{TIC} = \ln |\hat{\Sigma}| + 2\hat{\eta}/T,$$

where

$$\hat{\eta} = \sum_{t=1}^T \hat{\varepsilon}'_t \hat{\Sigma}^{-1} \hat{\varepsilon}_t h_t + \frac{1}{2} \left[T^{-1} \sum_{t=1}^T \left(\hat{\varepsilon}'_t \hat{\Sigma}^{-1} \hat{\varepsilon}_t \right)^2 - n(n+2) \right],$$

$h_t = \mathbf{X}'_t \left(\sum_{t=1}^T \mathbf{X}_t \mathbf{X}'_t \right)^{-1} \mathbf{X}_t$, $\hat{\varepsilon}_t$ are the OLS residuals, and $\hat{\Sigma}$ is the residual covariance matrix (see Yanagihara, 2006, for further details).

Normal	$T = 100$				$T = 250$			
	BIC	HQ	AIC	TIC	BIC	HQ	AIC	TIC
$p = 3$	26.07	56.01	67.30	37.56	82.17	96.15	86.98	94.38
$p < 3$	73.83	42.01	17.39	61.96	17.80	2.68	0.26	5.16
$t(10)$	$T = 100$				$T = 250$			
	BIC	HQ	AIC	TIC	BIC	HQ	AIC	TIC
$p = 3$	26.38	55.38	68.00	38.07	81.67	96.25	86.57	93.67
$p < 3$	73.55	42.69	17.33	61.47	18.32	2.69	0.26	5.78
$t(3)$	$T = 100$				$T = 250$			
	BIC	HQ	AIC	TIC	BIC	HQ	AIC	TIC
$p = 3$	27.25	55.29	67.03	37.20	80.22	95.05	86.49	81.89
$p < 3$	72.55	42.82	18.20	59.35	19.63	3.67	0.62	16.24
$t(1)$	$T = 100$				$T = 250$			
	BIC	HQ	AIC	TIC	BIC	HQ	AIC	TIC
$p = 3$	41.77	55.46	61.44	26.29	71.21	78.76	77.80	35.28
$p < 3$	53.87	35.27	19.83	57.15	24.76	13.25	6.62	53.60
Laplace	$T = 100$				$T = 250$			
	BIC	HQ	AIC	TIC	BIC	HQ	AIC	TIC
$p = 3$	29.03	57.88	69.15	39.87	82.56	96.10	87.07	92.21
$p < 3$	70.91	40.30	16.20	55.55	17.44	2.81	0.30	7.22

Table 5.1: Frequencies with which ICs find the pseudo-causal VAR(3) for a MVAR(1,2) DGP

cus on common cyclical feature test statistics using the tests reviewed above. For this study, we restrict our attention to two different data generating processes: a purely causal and purely noncausal VAR. More precisely, we consider a noncausal bivariate MVAR(0,2) with a reduced rank in coefficient matrices such as (5.14). For the causal MVAR(2,0) we have the same parameters in the first two lags, i.e., $\mathbf{\Pi}_1 = \mathbf{\Phi}_1$ and $\mathbf{\Pi}_2 = \mathbf{\Phi}_2$. Under the null hypothesis, there is a CCF [NCCF] relationship with a cofeature vector $\delta' = [1, -1]$. It should be noted that although we have strong co-movements between series, it is impossible to figure out graphically whether we have a causal or a noncausal pattern. To illustrate this point, Figures 1 and 2 compare causal and noncausal VAR(2) models with SCCF restrictions for different distributions.

We can now report the frequencies with which the three common feature test statistics that we propose reject the null that a combination of series is orthogonal to the past for the causal model or to the future for the noncausal model. At a 5% significance level the null follows a $\chi^2(3)$ under normality for a VAR(2) with one SCCF vector. An entry in the vicinity of 5% denotes an absence of size distortion.

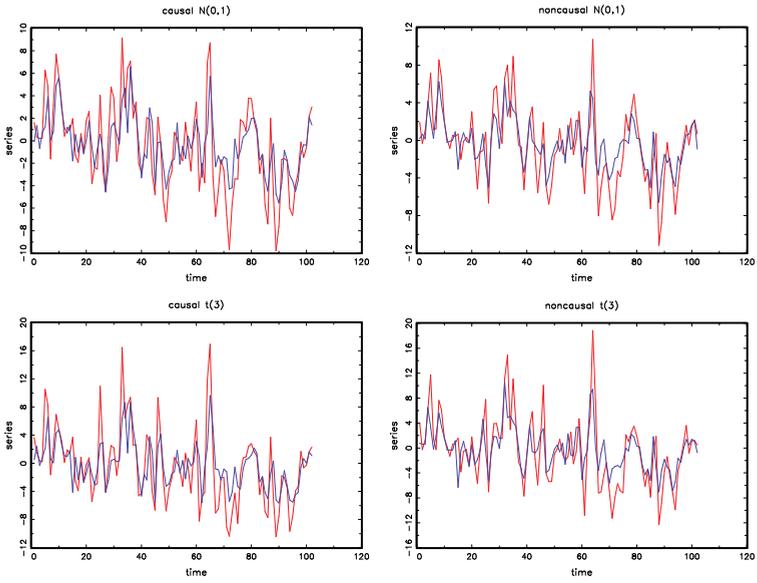


Figure 5.2: Causal and noncausal VAR(2) with SCCF - $N(0,1)$ and $t(3)$

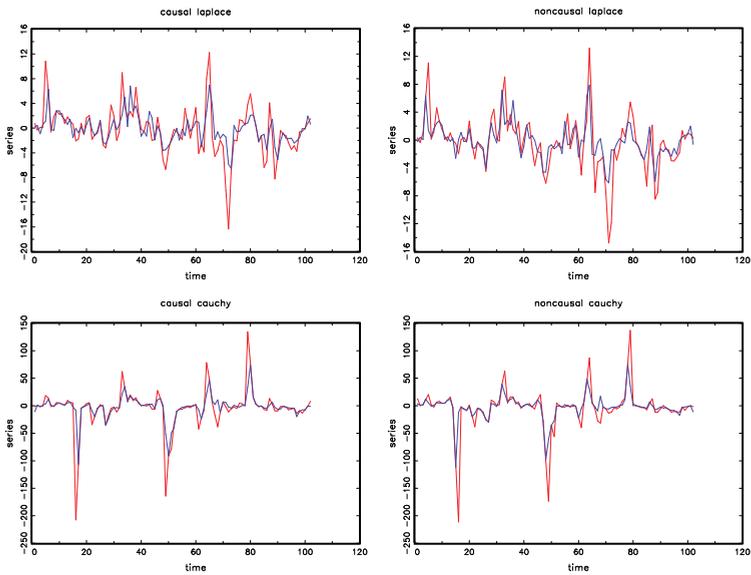


Figure 5.3: Causal and noncausal VAR(2) with SCCF - Cauchy and Laplace

$T = 100$						
DGP/instruments	$\{\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}\}$			$\{\mathbf{Y}_{t+1}, \mathbf{Y}_{t+2}\}$		
<i>Causal</i> VAR(2)	Cancor	GMM	GMM ^W	Cancor	GMM	GMM ^W
$N(0, 1)$	5.64	4.91	4.63	23.52	29.45	26.64
$t(10)$	5.37	4.82	4.58	23.47	29.52	25.40
$t(3)$	5.50	5.02	3.80	21.55	27.77	18.13
<i>Cauchy</i>	5.68	5.42	0.97	23.06	29.38	5.38
<i>Laplace</i>	5.83	5.10	4.30	27.67	38.70	31.51

$T = 500$						
	$\{\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}\}$			$\{\mathbf{Y}_{t+1}, \mathbf{Y}_{t+2}\}$		
<i>Causal</i> VAR(2)	Cancor	GMM	GMM ^W	Cancor	GMM	GMM ^W
$N(0, 1)$	5.16	5.08	4.56	91.84	92.54	91.99
$t(10)$	4.98	4.84	4.78	91.90	92.54	91.41
$t(3)$	4.84	4.72	4.15	89.17	89.89	78.32
<i>Cauchy</i>	4.76	4.73	1.00	67.46	69.22	12.61
<i>Laplace</i>	5.27	5.20	4.98	95.62	96.36	95.18

Table 5.2: Empirical size/power for the causal VAR(2) DGP

Results are based on 10,000 replications. Tables 2 and 3 reveal that, regardless of the distribution chosen to simulate processes, both canonical correlation and GMM tests do not suffer from severe size distortions. To obtain this result the “correct”¹⁰ set of instruments must be used when performing those tests, namely orthogonality conditions to $\{\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}\}$ [$\{\mathbf{Y}_{t+1}, \mathbf{Y}_{t+2}\}$] for the causal [noncausal] model. This result is very encouraging because it means that purely noncausal common features can be easily found by reverting the timeline of the instruments. The only test that faces a problem is the robust GMM when the Cauchy distribution is implemented. This is not surprising because in a heteroscedasticity robust correction à la White, it is assumed that the unconditional variance exists. In practice, the squares of the residuals cause problems in the Cauchy case when large “outlying” observations are introduced via the $\mathbf{B} = \text{diag}(u_1^2, u_2^2, \dots, u_T^2)$ matrix.

When the wrong set of instruments is used, namely only past variables for the noncausal model or the future variables for the causal specification, the presence of co-movements measured by a reduced rank in the dynamics of the systems is rejected. The rejection frequencies increase with the sample size, indicating that we are under the alternative indeed. Because we cannot distinguish whether co-movements are present in a causal or a noncausal framework by solely graphical

¹⁰Note that we solely mean lags for the causal model and leads for the noncausal model. We do not comment on under- or overspecification of these lag and lead orders.

$T = 100$						
DGP/instruments	$\{\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}\}$			$\{\mathbf{Y}_{t+1}, \mathbf{Y}_{t+2}\}$		
<i>Noncausal</i> VAR(2)	Cancel	GMM	GMM ^W	Cancel	GMM	GMM ^W
$N(0, 1)$	22.92	28.79	25.74	5.58	5.00	4.73
$t(10)$	22.86	28.88	24.50	5.74	5.05	4.46
$t(3)$	22.06	28.39	18.39	5.39	4.83	3.62
<i>Cauchy</i>	23.10	29.86	5.20	5.65	5.47	1.01
<i>Laplace</i>	27.02	38.69	30.94	5.45	4.86	4.23

$T = 500$						
	$\{\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}\}$			$\{\mathbf{Y}_{t+1}, \mathbf{Y}_{t+2}\}$		
<i>Noncausal</i> VAR(2)	Cancel	GMM	GMM ^W	Cancel	GMM	GMM ^W
$N(0, 1)$	91.39	92.06	89.55	5.13	4.94	5.02
$t(10)$	91.16	91.81	91.07	5.01	4.88	4.28
$t(3)$	89.64	90.43	78.77	5.11	4.98	4.17
<i>Cauchy</i>	67.93	69.55	12.72	4.55	4.53	1.00
<i>Laplace</i>	95.72	96.43	95.31	4.99	4.86	4.63

Table 5.3: Empirical size/power for the noncausal VAR(2) DGP

means, we recommend to first determine the VAR order p (potentially the pseudo-causal model) in empirical works and then to apply the common cyclical feature test statistics in both directions.

As an additional identification tool, we notice that one should reject the null hypothesis that the errors are *iid* if a pseudo-causal model is estimated when the data are generated by a noncausal model. Consequently, we can run the regression

$$\hat{\varepsilon}_t = \boldsymbol{\mu} + \boldsymbol{\Xi} \hat{\varepsilon}_{t-1}^2 + \mathbf{u}_t,$$

where $\hat{\varepsilon}_t$ are the residuals from the pseudo-causal VAR(p) model, and test for the null hypothesis $H_0 : \text{vec}(\boldsymbol{\Xi}) = \mathbf{0}$. We investigate the behavior of a Wald test that is based on the statistic

$$\hat{\chi}_W^{iid} = \text{vec}(\hat{\boldsymbol{\Xi}})' (\hat{\boldsymbol{\Sigma}}_{uu} \otimes (\mathbf{W}'\mathbf{W})^{-1})^{-1} \text{vec}(\hat{\boldsymbol{\Xi}}),$$

where $\hat{\boldsymbol{\Sigma}}_{uu}$ is the sample covariance matrix of the disturbance term \mathbf{u}_t and \mathbf{W} is the matrix of observations of the demeaned regressor $\hat{\varepsilon}_{t-1}^2$. We can also consider a LR test, which is based on the canonical correlations between $\hat{\varepsilon}_t$ and $\hat{\varepsilon}_{t-1}^2$. The

LR test statistic is given by

$$\xi_{LR}^{iid} = -T \sum_{i=1}^m \ln(1 - \hat{\lambda}_i),$$

where $\hat{\lambda}_i$ is the i -th smallest squared canonical correlation coming from the quantity $\text{CanCor} \{ \hat{\varepsilon}_t, \hat{\varepsilon}_{t-1}^2 \mid 1 \}$. Under the null hypothesis, both ξ_W^{iid} and ξ_{LR}^{iid} are asymptotically distributed as $\chi^2(n^2)$.

In order to evaluate the extent to which the above tests are able to determine whether we have a causal or a noncausal VAR, we use the same DGPs (either causal and noncausal) as before, then we estimate a (pseudo-)causal VAR(2) model and use both the statistics ξ_W^{iid} and ξ_{LR}^{iid} . The results can be found in Table 5.4. We obtain the well-known result that, under Gaussianity, causal and noncausal models cannot be distinguished. The rejection frequencies are in the vicinity of 5% for both sample sizes and both tests. When the model is causal we do not reject the null of *iid*-ness using both the Wald and the LR approaches. The rejection frequencies are all around 5% regardless of the distribution considered. The only exception is the Cauchy distribution which is slightly undersized. It is likely that other critical values should be considered in this case. When the true process is a noncausal VAR, both the Wald and the LR are able to reject the *iid* null hypothesis in large samples when the departure from the Normal distribution is large. The (non-size-adjusted) power is pretty low for $T = 100$, however with rejection frequencies of about 25% and around 33% with $T = 500$ for the Laplace. The $t(10)$ induces low power, which is expected as this specification is the closest to Gaussianity. The power is relatively high for the Cauchy and the $t(3)$. Alternative tests are worth investigating, but this is out of the scope of this chapter.

5.5 Empirical Illustrations

5.5.1 Price Indicators

We first consider 20 price series from “The Department Store Inventory Price Index” dataset. Those data are based on inventory weighted price indices of goods carried by department stores. The Bureau of Labor Statistics (BLS) maintained these statistics until the end of 2013. We work with monthly data from January 1980 to December 2013. Indices are available for the following 23 items: Piece goods, Domestic and draperies, Women’s and children’s shoes, Men’s shoes,

	$T = 100$		$T = 500$	
<i>Causal</i> VAR(2)	<i>Wald</i>	<i>LR</i>	<i>Wald</i>	<i>LR</i>
$N(0, 1)$	6.23	5.50	5.18	5.08
$t(10)$	5.89	5.16	5.14	4.96
$t(3)$	5.68	5.14	5.32	5.22
<i>Cauchy</i>	4.76	4.52	3.25	3.22
<i>Laplace</i>	5.66	5.05	5.51	5.29

	$T = 100$		$T = 500$	
<i>Noncausal</i> VAR(2)	<i>Wald</i>	<i>LR</i>	<i>Wald</i>	<i>LR</i>
$N(0, 1)$	5.71	4.85	5.14	5.00
$t(10)$	10.68	9.56	13.22	12.85
$t(3)$	42.67	40.58	75.35	75.13
<i>Cauchy</i>	81.20	80.20	98.57	98.56
<i>Laplace</i>	25.24	23.57	33.51	33.04

Table 5.4: Empirical size/power for the causal vs. noncausal VARs

Infants' wear and furniture, Women's underwear, Women's and girls' hosiery, Women's and girls' accessories, Women's outerwear & girls' wear, Men's clothing, Men's furnishings, Boys' wear, Jewelry and silverware, Notions, Toilet articles & drugs, Furniture and bedding, Floor coverings, Housewares, Major appliances, Radios and television sets. We do not include the last three items of the dataset in our analysis (Recreation & education, Home improvements, Automotive accessories) because those series only start in 1986.

Common cyclical feature test statistics have severe size distortions and low power when a wrong model for the seasonality is considered (Hecq, 1998; Cubadda, 1999). We refrain from using standard seasonal adjustment methods such as X13-ARIMA and TRAMO/SEATS, as they directly affect (partial) autocorrelation functions which might create spurious causal and noncausal dynamics (see Hecq et al., 2017c). We assume for this application that there is only one unit root at the zero frequency and that seasonality is deterministic. This means that we take the monthly growth rates $\Delta \ln P_{i,t}$ ($i = 1, \dots, 20$) in deviation from twelve deterministic monthly dummy variables. We construct 190 bivariate models from the 20 residual series. We have the following lag lengths for each of the 190 pairs of VARs: $p = 0$ (31%), $p = 1$ (5%), $p = 2$ (37%), $p = 3$ (5%), $p \geq 4$ (21%). There are consequently 59 pairs that are multivariate white noise and for which we should find both a causal and a noncausal common feature relationship. We use $p = 4$ and test for CCF and NCCF using the LR test. We do not reject the null of 74 causal

common feature relationships and 88 noncausal ones. This means that on the top of 59 multivariate white noise we have 15 additional causal SCCF relationship and 30 noncausal relationships. The impact of the assumption of stochastic versus deterministic seasonality is an important issue to investigate. We do not identify mixed models as we have seen (Section 5.2.4) that we would not have detected the presence of common cycles if the processes were mixed causal-noncausal.

5.5.2 Commodity Prices Co-Movements

For the second application we consider raw monthly commodity prices observed during the period January 1992 to October 2016, i.e., 298 observations, released by the IMF.¹¹ These are benchmark prices which are representative of the global market. They are determined by the largest exporter of a given commodity. IMF releases many different individual commodity prices but we only focus on the following three indices:

- RAWM: Agricultural Raw Materials Index, 2005 = 100, includes Timber, Cotton, Wool, Rubber, and Hides Price Indices. IMF name: PRAWM;
- GAS: Natural Gas, Indonesian Liquefied Natural Gas in Japan, US Dollar per Million Metric British Thermal Unit. IMF name: PNGASJP;
- OIL: Crude Oil (Petroleum), Price index, 2005 = 100, simple average of three spot prices; Dated Brent, West Texas Intermediate, and the Dubai Fateh. IMF name: POILAPSP.

We consider the vector of series

$$\mathbf{Y}_t = [\Delta \ln RAWM_t, \Delta \ln GAS_t, \Delta \ln OIL_t]',$$

where the monthly growth rates of the series are taken as we do not reject the zero frequency unit root in all three cases. The resulting series are displayed in Figure 5.4.

The pseudo-causal VAR order is detected to be $p = 12$ according to AIC in combination with cross correlogram analysis and seasonal dummies are included because they are jointly significant according to both ordinary (s.e._{ols}) and heteroscedasticity robust (s.e._{hcese}) F -tests. To look at the robustness of our results we

¹¹IMF Primary Commodity Prices, see <http://www.imf.org/external/np/res/commod/index.aspx>.

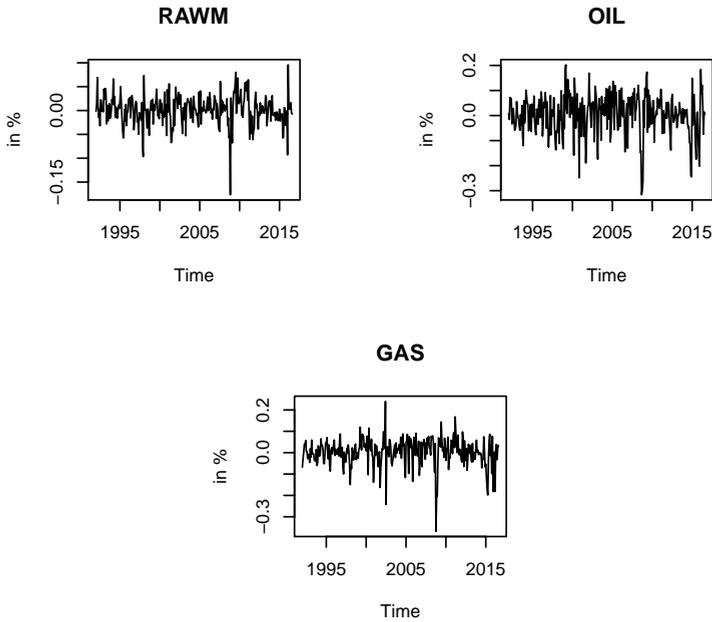


Figure 5.4: Growth rates of Raw Materials, Crude Oil and Natural Gas

have also considered Y_t^{SA} where series are taken in deviation from a regression on seasonal dummies. There are signs of non-normality according to the Jarque-Bera (JB) test. Table 5.5 displays p -values of the above tests.

Series	p	JB-test			F -test					
					s.e.ols	s.e.hcse	s.e.ols	s.e.hcse	s.e.ols	s.e.hcse
Y_t	12	0.000	0.000	0.151	0.000	0.001	0.769	0.898	0.329	0.426
Y_t^{SA}	12	0.000	0.000	0.124	-	-	-	-	-	-

Table 5.5: Summary statistics of the (pseudo-)causal VAR

It is clear from Table 5.6 that both the canonical correlation test ξ_{LR} and the robust GMM J_2 -test do not reject the null of a common feature vector in the noncausal framework while it is rejected (at 5%) in the purely causal VAR. Eigenvalues are also very interesting to investigate. In the causal VAR there is not much difference between the first two smallest eigenvalues. Therefore, it is difficult to claim that the first value is not different from zero while the second one

is. This is directly in contrast with the noncausal case as we do observe a large difference between them.

	CCF _{s=1}			NCCF _{s=1}		
	CanCor _{ξ_{LR}}	$\tilde{\lambda}_i$	GMM _{J₂}	CanCor _{ξ_{LR}}	$\tilde{\lambda}_i$	GMM _{J₂}
\mathbf{Y}_t	0.015	(0.172, 0.186, 0.513)	0.031	0.333	(0.121, 0.221, 0.567)	0.090
\mathbf{Y}_t^{SA}	0.044	(0.158, 0.196, 0.520)	0.013	0.716	(0.096, 0.272, 0.553)	0.363

Table 5.6: p -values of tests - causal vs. noncausal VARs

The presence of noncausal common features detected in both applications can imply that there exist some commonalities in the bubble behavior of several variables. It can also be the sign of collinear nonlinear impulse response functions as noncausal linear VARs have a nonlinear causal interpretation (Gourieroux and Zakoian, 2016).

5.6 Conclusion

This chapter provides a novel explanation for the inability of detecting common features in a VAR framework. We show that common features that cannot be detected in the causal (i.e., backward-looking) representation might be revealed in the noncausal (i.e., forward-looking) dynamics of the series. Whereas the difference between data generated from purely causal, purely noncausal and mixed causal-noncausal VARs is not always graphically visible, it is well-known that the latter two are able to generate much richer dynamics than the conventional causal VAR. This makes it possible to detect for example speculative bubbles and asymmetric cycles which are common to a set of series.

We propose tools to highlight the potential presence of noncausal co-movements. Using both sets of lag and lead instruments within a canonical correlation or a GMM framework, we show how additional relationships are discovered between series both in Monte Carlo simulations and two empirical illustrations. Interestingly, we find that the presence of a reduced rank structure in the dynamics of purely causal and noncausal systems with more than one lag/lead permits identification in the usual Gaussian framework. This result cannot be extended to mixed causal-noncausal models, as the instruments are correlated with the error term of the reduced rank model by construction. We identify common bubbles in both commodity prices and price indicators.

Chapter 6

The MARX Package in R

This chapter presents the **MARX** package for the analysis of mixed causal-noncausal autoregressive processes. The distinctive feature of MARX models is that they abandon the Gaussianity assumption on the error term. This deviation from the Box-Jenkins approach allows researchers to distinguish backward- (causal) and forward-looking (noncausal) stationary behavior in time series (see e.g. Hecq et al., 2016 for an overview). The **MARX** package offers functions to simulate, estimate and select mixed causal-noncausal autoregressive models, possibly including exogenous regressors. The procedures for this are discussed in Hecq et al. (2016) for the MAR, and Hecq et al. (2017a) for the MARX respectively.¹

¹This chapter is based on the paper Hecq, A., Lieb, L. and S. Telg (2017b), Simulation, Estimation and Selection of Mixed Causal-Noncausal Autoregressive Models: The **MARX** Package, *Working Paper*, available at SSRN: <https://ssrn.com/abstract=3015797>.

6.1 Introduction

Various scientific fields consider models where the present output is a function of current and future values of the input variable. Examples can be found in physics (see e.g., Schmidt, 1978), signal processing (see e.g., Jackson, 1989) and astronomy (e.g., Scargle, 1981). This phenomenon, known as *noncausality* or *acausality*, has only relatively recently received more attention in the econometric literature. In particular, the noncausal autoregressive model has been shown to provide many interesting opportunities from both a statistical and empirical point of view. Brockwell and Davis (1991) discuss the possibility of rewriting a process with explosive roots in direct time into a process with roots strictly outside the unit circle in reverse time. This means that except for the unit root case, a convergent solution can always be found either backward or forward in time. From an applied point of view, noncausal models are shown to improve forecast performances over conventional causal autoregressive process (Lanne et al., 2012a). By construction, a noncausal model is nonfundamental (Alessi et al., 2011) and is often found to be a stationary solution to rational expectation models (Gouriéroux et al., 2016). Another important feature of noncausal models is that they are able to replicate dynamics (e.g., bubbles and asymmetric cycles) which previously could only be obtained using highly nonlinear and complex models (Gouriéroux and Zakoïan, 2016). These properties lead to the interest in mixed causal-noncausal autoregressive (MAR) models, which have both a backward- and forward-looking component.

The **R** package **MARX** is developed in order to extend the conventional time series analysis to allow for a noncausal component. The existing autoregressive moving average (ARMA) modelling is based on the Box-Jenkins approach, which only considers a backward-looking filter. When the error term is assumed to be Gaussian, it is well-known (see e.g. Brockwell and Davis, 1991) that the backward- and forward-looking model is statistically indistinguishable. This is due to the fact that the Gaussian distribution is fully characterized by its second order properties. Autocovariances and spectral densities are fully symmetric and thus the same both backward and forward in time. As soon as the Gaussianity assumption on the error term is abandoned, one can discriminate between these processes which enlarges the set of potential models considered in a model selection procedure. An overview of the practical relevance of these models in the econometric literature

and an illustration of their implementation can be found in Hecq et al. (2016) and references therein.

Besides allowing for noncausality (i.e., MAR), the **MARX** package also offers the possibility to have various exogenous regressors (X) as introduced in Hecq et al. (2017a). To that end, the package contains an integrated set of **R** functions to do time series analysis in **R** using the MARX framework. The functionalities include (i) simulation, (ii) estimation, (iii) model selection and (iv) inference on mixed causal-noncausal processes. Some functions are written using **S3** classes and provide methods such as `summary()`, `coef()` and `resid()`. The package also offers an interface-based function that guides the user through the complete process of modelling mixed causal-noncausal processes. Various additional functions are available for users to perform certain steps manually. All implemented methods are discussed in this chapter and illustrated in an empirical application on commodity prices with the U.S. exchange rate as exogenous variable.

The outline of the chapter is as follows. Section 6.2 reviews mixed causal-noncausal autoregressive models and focuses on how they differ from the conventional (causal) autoregressive process using findings from Hecq et al. (2017a) and Lanne and Saikkonen (2011a). Section 6.3 introduces the **R** package **MARX** and illustrates how to simulate, estimate and perform model selection. Section 6.4 presents an application using data on commodity prices. Section 6.5 concludes.

6.2 The $\text{MARX}(r, s, q)$ Model

The introduction of noncausality in autoregressive processes with (possibly) exogenous regressors has important implications for the simulation, estimation and selection of these models. We comment on the necessity of a non-Gaussian error term and the decomposition of a mixed causal-noncausal process into causal and noncausal filtered values to accommodate the simulation of such paths. In this section, we first introduce the notation and present the mixed causal-noncausal autoregressive model with exogenous regressors in its most general form. We provide a detailed explanation on how to simulate MARX processes. Subsequently, we discuss the maximum likelihood estimation procedure (based on Student's t -likelihood) used to estimate these models. The section concludes with the two-step approach used to perform model selection. We introduce the notion of pseudo-causal or equivalently weak linear causal model.

6.2.1 Model Specification

Let y_t be the variable of interest which is observed over the time period $t = 1, \dots, T$. Let $x_{i,t}$ ($i = 1, \dots, q$) be the i th variable in a set of q strictly exogenous variables for y_t and $\beta \in \mathbb{R}^q$ a vector of parameters. Then we can define $\mathbf{X}_t = [x_{1,t}, \dots, x_{q,t}]' \in \mathbb{R}^q$ as the vector of all q exogenous variables at time t . The MARX(r, s, q) for a stationary time series y_t is given by²

$$\phi(L)\varphi(L^{-1})y_t - \beta' \mathbf{X}_t = \varepsilon_t, \quad (6.1)$$

where $\phi(L)$ is a lag polynomial of order r , $\varphi(L^{-1})$ a lead polynomial of order s and $r + s = p$. The error term ε_t is assumed to be *iid* non-Gaussian. When $\varphi_1 = \dots = \varphi_s = 0$, the process y_t is a purely causal autoregressive model with strictly exogenous variables, denoted MARX($r, 0, q$) or simply ARX(r, q):

$$\phi(L)y_t - \beta' \mathbf{X}_t = \varepsilon_t. \quad (6.2)$$

Specification (6.2) can be seen as the standard backward-looking ARX model. Conversely, the process in (6.1) reduces to a purely noncausal MARX($0, s, q$) given by

$$\varphi(L^{-1})y_t - \beta' \mathbf{X}_t = \varepsilon_t, \quad (6.3)$$

when $\phi_1 = \dots = \phi_r = 0$. The concepts of causality and noncausality are defined in terms of the strictly stationary solution of the model. To that end, we assume that both polynomials in (6.1) have their zeros outside the unit circle:

$$\phi(z) \neq 0 \text{ for } |z| \leq 1 \text{ and } \varphi(z) \neq 0 \text{ for } |z| \leq 1. \quad (6.4)$$

When $q = 0$, the process in (6.2) [(6.3)] reduces to a purely causal [noncausal] AR process that has a one-sided MA representation consisting of only past [future] and current values of ε_t . For the general process in (6.1) these conditions however imply that the process y_t follows a two-sided MA representation involving past, current and future values of ε_t . In case $q > 0$, the processes no longer have a

²We skip deterministic elements for the sake of representation.

strictly stationary solution solely in terms of ε_t , but involve both \mathbf{X}_t and ε_t :

$$y_t = \pi(L, L^{-1}) [\varepsilon_t + \beta' \mathbf{X}_t] = \sum_{j=-\infty}^{\infty} \pi_j z_{t-j}, \quad (6.5)$$

with $z_{t-j} = \varepsilon_{t-j} + \sum_{i=1}^q \beta_i x_{i,t-j}$ and $\pi(z, z^{-1}) = [\phi(z)\varphi(z^{-1})]^{-1}$. Similar to Gouriéroux and Jasiak (2016), we note that the polynomials $\phi(z)$ and $\varphi(z^{-1})$ are invertible and their inverses create infinite series in z and z^{-1} respectively, causing (6.5) to hold almost surely (see Brockwell and Davis, 1991, proposition 13.3.1 for more details).

6.2.2 Simulation

From (6.1), Hecq et al. (2017a) construct the filtered values (u, v) which decompose the MARX process in terms of its unobserved causal and noncausal component:

$$u_t \equiv \phi(L)y_t \leftrightarrow \varphi(L^{-1})u_t - \beta' \mathbf{X}_t = \varepsilon_t, \quad (6.6)$$

$$v_t \equiv \varphi(L^{-1})y_t \leftrightarrow \phi(L)v_t - \beta' \mathbf{X}_t = \varepsilon_t. \quad (6.7)$$

These filtered values establish a deterministic dynamic relationship between the unobserved components (u_t, v_t) , the exogenous variables \mathbf{X}_t and the process y_t , which can be used to simulate various MARX(r, s, q) series. The main difficulty for generating MARX(r, s, q) is that the process depends on both the past and the future: simulating such a process requires both initial and terminal values to enter the model simultaneously. Filtered values are used to circumvent this problem. Defining $[\varphi(z^{-1})]^{-1} \equiv \delta(z^{-1})$, we can rewrite the second equality in (6.6) in the following way:

$$u_t = \sum_{j=0}^{\infty} \delta_j \left(\sum_{i=1}^q \beta_i x_{i,t+j} + \varepsilon_{t+j} \right) = \sum_{j=0}^{\infty} \delta_j z_{t+j}. \quad (6.8)$$

In a similar fashion, when we take $[\phi(z)]^{-1} \equiv \alpha(z)$, we obtain for v_t :

$$v_t = \sum_{j=0}^{\infty} \alpha_j \left(\sum_{i=1}^q \beta_i x_{i,t-j} + \varepsilon_{t-j} \right) = \sum_{j=0}^{\infty} \alpha_j z_{t-j}. \quad (6.9)$$

Using these expressions, MARX(r, s, q) can be constructed directly by means of the definitions given in either (6.6) or (6.7). That is, the causal and noncausal

components (u_t, v_t) can be simulated independently and can be interpreted as a causal [noncausal] “error term” of a purely noncausal [causal] autoregression.

We characterize the simulation steps using (6.6). Note that the first equality $\phi(L)y_t = u_t$ appears like a conventional causal autoregressive process. In order to simulate such a process, one needs r starting values, say $y_{-1}^*, \dots, y_{-r}^*$ to create the first value y_1 . Additionally, one needs the value v_1 which is usually a draw from a desired distribution in the case of a causal autoregressive process. In the MARX case, however, v_1 is represented as a linear combination of current and future values of ε_t and \mathbf{X}_t as can be seen in (6.8). If we consider a truncation m sufficiently large for the infinite sum, v_1 can be constructed by simulating long paths of ε_t and \mathbf{X}_t such that z_1 up to z_{1+m} are available. Now that y_1 is generated, it can be used to construct the next value y_2 . In general, the following two steps can be used to simulate an MARX processes based on (6.6):

1. Generate a path of length $(T + m)$ for ε_t and \mathbf{X}_t and simulate the values of u_t using a truncated version of (6.8).
2. Create *starting* values y_{-1}^* up to y_{-r}^* and simulate the process y_t like a conventional causal autoregressive process.

It is also possible to base the simulation on (6.7), then one typically generates the series “backwards”. In that case, the following steps are needed:

1. Generate a path of length $(T + m)$ for ε_t and \mathbf{X}_t and simulate the values of v_t using a truncated version of (6.9).
2. Create *terminal* values y_{T+1}^* up to y_{T+s}^* and simulate the process y_t like a noncausal autoregressive process (see e.g., Gouriéroux and Jasiak, 2016).

Example 6.1. An MARX(1,1,1) can be simulated according to

$$y_t = \varphi_1 y_{t+1} + \sum_{j=0}^{\infty} \phi^j (\beta_1 x_{1,t-j} + \varepsilon_{t-j}), \quad (6.10)$$

where a truncation m sufficiently large has to be considered for the infinite sum.³ The simulation of this process consists of two steps. Firstly, simulate a long path of length $T + m$ for $x_{1,t}$ and ε_t . Accordingly, one can construct the sequence v_t

³Since $\deg(\phi(z)) = 1$ in this instance, it is straightforward to compute the inverse of this polynomial. For more complicated polynomials, one could compute a companion matrix to find its inverse.

for $t = 1, \dots, T$. Secondly, choose a terminal value, say y_{T+1}^* . Using (6.10), we are going to simulate the series “backwards”. That is, to simulate y_T , we put the terminal value y_{T+1}^* for y_{t+1} in (6.10) and use the value v_T simulated in the first step. Now that y_T becomes available, it can be used in combination with v_{T-1} to construct y_{T-1} . We continue this procedure until we reach y_1 .

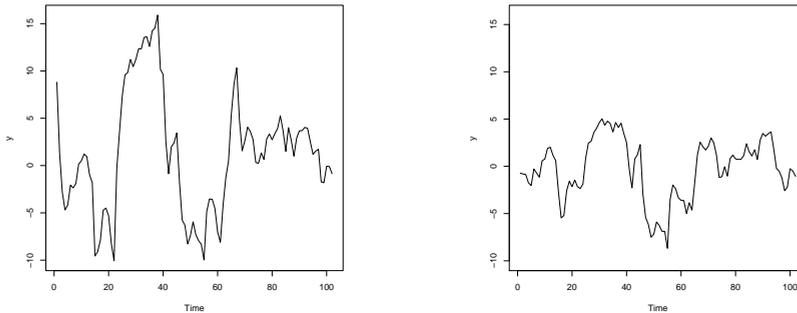


Figure 6.1: A simulated MARX process (left) and the same process without exogenous variable (right)

Figure 6.1 shows simulated paths of an MARX(1,1,1) and MAR(1,1) process with $[\phi_1, \varphi_1]' = [0.3, 0.9]'$, $\beta_1 = 0.3$, $x_{1,t} \stackrel{iid}{\sim} t(1, 0)$ and $\varepsilon_t \stackrel{iid}{\sim} t(3, 0)$ with $t(\nu, \sigma)$ being the Student’s t distribution with degrees of freedom parameter ν and scale parameter σ . The truncation m is set at 10,000. It can be seen that both processes generally move similarly with the major exception that the MARX process contains more peaks and troughs, which are also more extreme in comparison. This is due to the choice of $x_{1,t}$, which is chosen to be standard Cauchy distributed for expository purposes. Hence, the MARX specification takes into account shocks that cannot be explained by past, current and future values of the dependent variable, but which are present because of major changes in explanatory exogenous variables at some specific points in time.

6.2.3 Maximum Likelihood Estimation

For MAR models, the non-Gaussianity assumption on the error term allows for the discrimination between purely causal, purely noncausal and mixed causal-noncausal autoregressive processes (Lanne and Saikkonen, 2011). The reason for this is that the Gaussian distribution is the only distribution for which the complete

probability structure is captured in first- and second-order moments. Since the autocovariance function is symmetric, considering p lags is completely similar to considering p leads or a combination of r lags and s leads with $r + s = p$. Hence, in order to distinguish between these processes, we need an estimation method which is based on more information than solely second-order properties.⁴

Lanne and Saikkonen (2011a) propose the approximate maximum likelihood (AML) estimator based on non-Gaussian error distributions satisfying regularity conditions in Andrews et al. (2006). In empirical applications, the t -distribution (with scale parameter) is often employed, as it offers the flexibility of capturing leptokurtic and skewed processes (Lanne and Saikkonen, 2011a; Hecq et al., 2016). Hecq et al. (2017a) extend these results to mixed autoregressive processes including exogenous regressors. Let $\mathbf{y} = [y_1, \dots, y_T]'$, $n \equiv (T-p)$ and $\boldsymbol{\lambda} = [\sigma, \nu]'$. If ε_t is a sequence of *iid* t -distributed zero mean random variables, then the corresponding approximate log-likelihood function is given by

$$l_y(\boldsymbol{\phi}, \boldsymbol{\varphi}, \boldsymbol{\beta}, \alpha, \boldsymbol{\lambda} | \mathbf{y}) = n \left[\ln(\Gamma((\nu + 1)/2)) - \ln(\sqrt{\nu\pi\sigma^2}) - \ln(\Gamma(\nu/2)) \right] \\ - \frac{\nu + 1}{2} \sum_{t=r+1}^{T-s} \ln \left(1 + \frac{1}{\nu} \left(\frac{\boldsymbol{\phi}(L)\boldsymbol{\varphi}(L^{-1})y_t - \boldsymbol{\beta}'\mathbf{X}_t - \alpha}{\sigma} \right)^2 \right),$$

where an intercept α is introduced in model (6.1). The scale parameter is denoted by σ , the degrees of freedom by ν (> 0) and $\Gamma(\cdot)$ is the Gamma function. Thus, the AML estimator corresponds to the solution of the problem:

$$\hat{\boldsymbol{\theta}}_{ML} = \arg \max_{\boldsymbol{\theta} \in \Theta} l_y(\boldsymbol{\theta} | \mathbf{y}),$$

with $\boldsymbol{\theta} = [\boldsymbol{\phi}', \boldsymbol{\varphi}', \boldsymbol{\beta}', \alpha, \boldsymbol{\lambda}']'$ and Θ is a permissible parameter space containing the true value of $\boldsymbol{\theta}$, say $\boldsymbol{\theta}_0$, as an interior point. Hecq et al. (2017a) show that the asymptotic distribution of the regressor coefficients can be computed analytically, offering the possibility to compute corresponding standard errors in closed form. Define the $(n \times 1)$ series $\mathbf{u} \equiv \mathbf{U}_t^* = [u_{r+1}, \dots, u_{T-s}]'$ up to $\mathbf{U}_{t+s}^* = [u_{r+s+1}, \dots, u_T]'$, $\mathbf{V}_{t-r}^* = [v_1, \dots, v_{T-p}]'$ up to $\mathbf{v} \equiv \mathbf{V}_t^* = [v_{r+1}, \dots, v_{T-s}]'$, $\mathbf{X}_{i,t} = [x_{i,r+1}, \dots, x_{i,T-s}]'$ and $\boldsymbol{\varepsilon} = [\varepsilon_{r+1}, \dots, \varepsilon_{T-s}]'$. Then we construct the matrices $\mathbf{Z} = [\mathbf{U}_{t+1}^*, \dots, \mathbf{U}_{t+s}^*, \mathbf{X}_{1,t}, \dots, \mathbf{X}_{q,t}]$ and $\mathbf{Q} = [\mathbf{V}_{t-1}^*, \dots, \mathbf{V}_{t-r}^*, \mathbf{X}_{1,t}, \dots, \mathbf{X}_{q,t}]$,

⁴Hecq et al. (2017a) show that purely causal and purely noncausal ARX models can be distinguished using estimation based on second-order properties (OLS and Gaussian MLE). As this result does not extend to mixed processes, we however also consider an alternative estimation method (t -MLE) for MARX models.

which are of dimensions $(n \times (s + q))$ and $(n \times (r + q))$ respectively. Using this notation, we can write the autoregressions defined in (6.6) and (6.7) in matrix notation as follows:

$$\mathbf{u} = \mathbf{Z}\boldsymbol{\zeta} + \boldsymbol{\varepsilon},$$

$$\mathbf{v} = \mathbf{Q}\boldsymbol{\xi} + \boldsymbol{\varepsilon},$$

with $\boldsymbol{\zeta} = [\boldsymbol{\varphi}', \boldsymbol{\beta}']' \in \mathbb{R}^{s+q}$ and $\boldsymbol{\xi} = [\boldsymbol{\phi}', \boldsymbol{\beta}']' \in \mathbb{R}^{r+q}$.

Then, conditional on the unobserved causal and noncausal components discussed above, it can be shown that in the case of an MARX(r, s, q) model

$$\begin{aligned} \sqrt{n}(\boldsymbol{\zeta}_{ML} - \boldsymbol{\zeta}_0) &\xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \frac{\nu + 3}{\nu + 1} \sigma^2 \boldsymbol{\Upsilon}_{\boldsymbol{\phi}}^{-1}\right), \\ \sqrt{n}(\boldsymbol{\xi}_{ML} - \boldsymbol{\xi}_0) &\xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \frac{\nu + 3}{\nu + 1} \sigma^2 \boldsymbol{\Upsilon}_{\boldsymbol{\varphi}}^{-1}\right), \end{aligned}$$

holds. We use the notation $\boldsymbol{\Upsilon}_{\boldsymbol{\varphi}} = \mathbb{E}[\mathbf{Q}'\mathbf{Q}]$ and $\boldsymbol{\Upsilon}_{\boldsymbol{\phi}} = \mathbb{E}[\mathbf{Z}'\mathbf{Z}]$, where $\boldsymbol{\varphi}$ and $\boldsymbol{\phi}$ signify the relation between the unobserved values \mathbf{u}, \mathbf{v} and \mathbf{y} as defined in (6.6)-(6.7). These quantities can be estimated consistently by $(1/n) \sum_{i=1}^n \mathbf{Q}'_i \mathbf{Q}_i$ and $(1/n) \sum_{i=1}^n \mathbf{Z}'_i \mathbf{Z}_i$, where \mathbf{Q}_i [resp. \mathbf{Z}_i] denotes the i th row of the matrix \mathbf{Q} [resp. \mathbf{Z}]. For large ν , i.e., $\nu \rightarrow \infty$, l_y approaches the Gaussian (log)-likelihood, and the model parameters cannot be consistently estimated anymore.

6.2.4 Modelling Procedure

Lanne and Saikkonen (2011a) propose a two-step approach to identify MAR models; a procedure that got extended to MARX models in Hecq et al. (2017a). In the proposed model selection procedure, it is assumed that the exogenous variables are chosen a priori. The two-step approach can be summarized as follows:

1. Estimate by OLS conventional (causal) ARX models up to order p_{max} to determine the autoregressive order p that deletes all serial correlation from the residuals. This can be done using information criteria (e.g., AIC, BIC, HQ) in addition to diagnostic tests (e.g., Ljung-Box, correlogram). Tests are performed on the residuals of the chosen ARX(p, q) model, in particular tests for normality and *iid*-ness.
2. After p is established, select that MAR(r, s, q) with $r+s = p$ which maximizes the log-likelihood at the estimated parameters.

The model in step 1 is referred to as pseudo-causal model in Hecq et al. (2017a) and weak linear causal model in Gouriéroux and Zakoian (2016). The name ‘weak linear causal’ stems from the fact that, in the non-Gaussian world, there can only be one convergent linear representation which consists of *iid* noise. Hence, if the true data generating process is a linear noncausal model, a linear causal model⁵ has uncorrelated but not independent noise. Similarly, the name ‘pseudo-causal’ indicates that it is an existing representation, but not necessarily the linear representation with strong white noise. However, all of this only holds if we assume the existence of a linear representation with *iid* non-Gaussian error term. For this reason, the residuals of the pseudo-causal model chosen in Step 1 are tested against normality and *iid*-ness. Only if these two null hypotheses are rejected, it makes sense to perform Step 2.⁶ In this step, we estimate $p + 1$ models, evaluate the log-likelihood of those models at the estimated parameters and select the model with the highest value.

6.3 The R Package **MARX**

The R package **MARX** offers a straightforward way to deal with MARX models. Most functions are constructed in such a way that they directly offer the possibility to perform data analysis. Complementary to that, there is also a function that allows to simulate MARX processes. In particular, functions for (i) simulation, (ii) estimation, (iii) model selection and (iv) inference are available for univariate time series.⁷ We discuss them in this section and assign them to procedures in order to structure the package.

6.3.1 Structure

The structure of the **MARX** package is depicted in Figure 6.2. The two main building blocks of the package are ‘Simulation’ and ‘Model selection’. The first one only consists of one function `sim.marx()` and can be used to simulate various MARX processes. The second block ‘Model selection’ can be divided into two sub-groups again: ‘Pseudo-causal model’ and ‘Final MARX model’. As explained in

⁵The Wold Representation Theorem states that every covariance-stationary sequence has a representation as a deterministic and stochastic part, where the latter is a linear combination of past and current weak white noise.

⁶If normality cannot be rejected, purely causal, noncausal and mixed causal-noncausal representations cannot be distinguished. If *iid*-ness cannot be rejected, it means that we already found the only representation with *iid* noise and thus we can stop.

⁷We developed similar routines in `matlab` for MAR models, which are available upon request.

Section 6.2.4, the model selection procedure for MARX processes follows a two-step approach. Hence, the blocks ‘Pseudo-causal model’ and ‘Final MARX model’ contain the estimation and selection of purely causal ARX and mixed causal-noncausal ARX models respectively. The functions that are employed in these stages can also be found in Figure 6.2. It is important to note that also S3-functions `pseudo()` and `mixed()` are available. These functions are programmed such that their output closely resembles the class `lm` in `R`. As such, standard functions like `fitted.values()`, `resid()` and `summary()` can be applied to these objects. The package `MARX` has one more function available which is not depicted in Figure 6.2, namely the `marx()` function. It is set up to be interface-based and user-friendly, as it guides the user through the complete model selection procedure. This function is discussed in more detail in the empirical application in order to show the practical relevance and the straightforward usage of the `MARX` package even further.

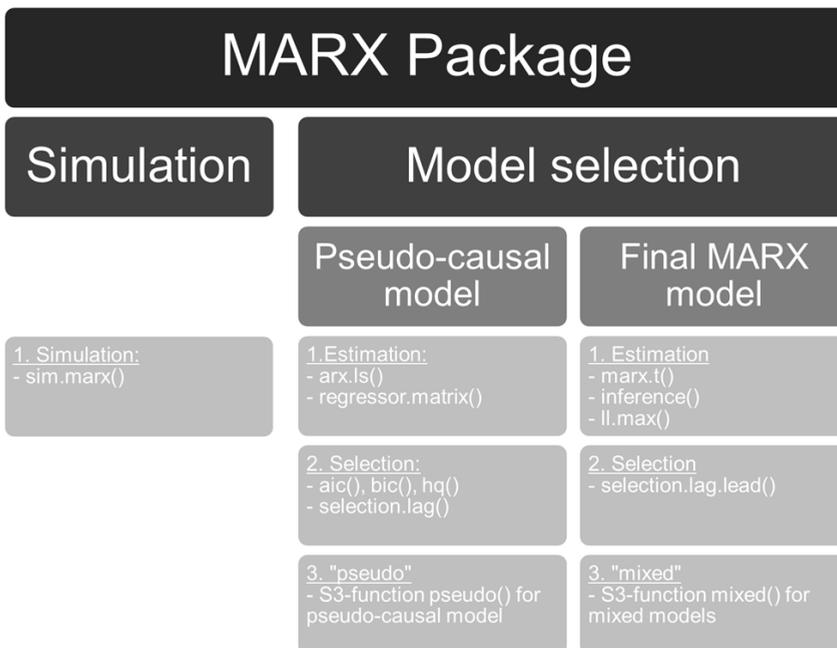


Figure 6.2: Structure of the MARX package

6.3.2 Simulation

The function `sim.marx()` can be used to simulate MARX processes easily. In order to do so, the user has to specify six elements. The first two are the distributions of the error term (`dist.eps`) and the exogenous variable (`dist.x`). These quantities have to be specified as vectors containing the distribution between quotation marks and the corresponding parameters. The user can choose from three different distributions: the rescaled Student's t -distribution, the normal distribution or the stable distribution.⁸ The third argument of the `sim.marx()` function is the number of observations (`obs`), while the last three arguments are the vectors of causal and noncausal parameters (`c_par` and `nc_par`) and the parameter of the exogenous variable (`exo_par`).⁹

As an example, we simulate 200 observations of an $MARX(2,1,1)$ with $\phi = [0.2, 0.3]'$, $\varphi_1 = 0.7$ and $\beta_1 = 0.5$. The error term is assumed to follow a t -distribution with 3 degrees of freedom and scale parameter equal to 0, while the exogenous variable follows the same distribution however with degrees of freedom equal to 2 and scale parameter 1. We can use the following line in **R** to simulate such a process:

```
> sim.data <- sim.marx(c('t',3,0),c('t',2,1),200,c(0.2,0.3),0.7,0.5)
```

The output of this function, `sim.data`, is a list of two elements: the simulated process y_t (accessible as `sim.data$y`) and the simulated exogenous variable x_t (accessible as `sim.data$x`).¹⁰ Using the `ts.plot` function from the `stats` package, we obtain the graphs from Figure 6.3. We will use these generated data, which we save as `y` and `x` respectively, to characterize the functions in the 'Model Selection' block of the package. Note that one can also easily simulate, estimate and select mixed causal-noncausal models without exogenous variables by setting `x` and `dist.x` equal to `NULL` or "not". That is, if we want to simulate the same mixed causal-noncausal process from the example, however this time without exogenous variable, we run the following code:

```
> sim.data <- sim.marx(c('t',3,0),NULL,200,c(0.2,0.3),0.7,0)
```

⁸Sequences of t - and normal distribution are simulated using the `stats` package; for a series of stable distributed random variables this is the `stabledist` package. Please refer to their documentation for more details.

⁹Note that one can only simulate $MARX(r, s, 1)$ processes.

¹⁰As we consider only one exogenous variable in the sequel of this chapter, we choose for the notation x_t rather than $x_{1,t}$ as used earlier.

Note that we filled in the value ‘0’ for `exo_par`, however this entry becomes irrelevant as soon as `dist.x` is set to `NULL` or “not”.

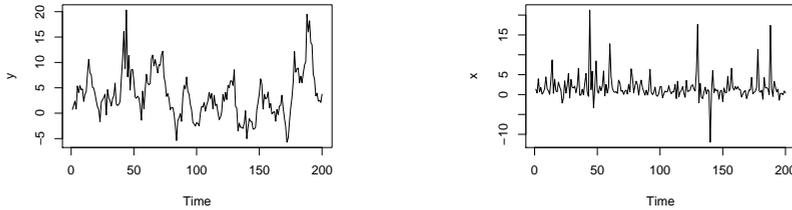


Figure 6.3: The simulated MARX process y_t (left) and the exogenous variable x_t (right)

6.3.3 Pseudo-Causal Models

Imagine we treat the series generated in section 6.3.2, as data y_t and x_t that we have at our disposal. Then the first step towards identifying mixed causal-noncausal autoregressive models is to identify the total autoregressive order p . As proposed in Hecq et al. (2017a), one can use information criteria (IC) like AIC (Akaike), BIC (Bayesian) and HQ (Hannan-Quinn) to determine this order. To that end, we need to estimate $ARX(p^*)$ models for $p^* = 0, \dots, p_{max}$ and compute the information criteria for all these models. One chooses the lag order p^* for which the information criteria is minimized to be the final p . The value p_{max} has to be selected by the user. The function `selection.lag()` is designed to compute exactly these values and pick their minimum. Let us take, for expository purposes, $p_{max} = 6$. Then we can call the function

```
> selection.lag(y,x,6)
```

which produces the following output:

```
Order Selection Criteria Pseudo Causal Model:
```

```
BAYESIAN INFORMATION CRITERION
```

```
      p = 0    p = 1    p = 2    p = 3    p = 4    p = 5    p = 6
[1,] 5.718606 4.455083 4.436168 4.32274 4.351089 4.366575 4.396754
```

Minimum value attained at $p = 3$

AKAIKE INFORMATION CRITERION

	$p = 0$	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$p = 6$
[1,]	5.685623	4.405435	4.369738	4.23941	4.250739	4.249083	4.261997

Minimum value attained at $p = 3$

HANNAN-QUINN INFORMATION CRITERION

	$p = 0$	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$p = 6$
[1,]	5.698971	4.425529	4.396627	4.273143	4.291365	4.296654	4.316564

Minimum value attained at $p = 3$

Since we chose $r = 2$ and $s = 1$ for the simulated process, the information criteria correctly indicate a total autoregressive order of 3. However, in real empirical applications, it is advisable to perform additional tests on the residuals to check whether all serial correlation has been treated. This can easily be done by creating an object of the class “pseudo”:

```
> pseudo.model <- pseudo(y,x,3)
```

The residuals are readily available using `pseudo.model$residuals` or `resid(pseudo.model)`. Various tests from e.g. the package `stats` can now be used to investigate these residuals. Using the `summary()` function on the object `pseudo.model`, we obtain

```
\$call
```

```
pseudo.default(y = y, x = x, p = 3)
```

```
\$coefficients
```

	Estimate	Std.Err	t value	p value
int	0.03082937	0.19317881	0.1595898	8.733720e-01
lag 1	0.71113979	0.05567727	12.7725323	1.927037e-27
lag 2	0.36245418	0.06332079	5.7240941	3.949486e-08
lag 3	-0.30451150	0.05552195	-5.4845244	1.295894e-07
exo 1	0.46051522	0.04604578	10.0012469	3.197321e-19

```
attr("class")
```

```
[1] "summary.pseudo"
```

Note that both `selection.lag()` and `pseudo()` make use of other functions that we did not discuss. The functions `aic()`, `bic()` and `hq()` can be used to compute the information criteria AIC, BIC and HQ separately; the function `ar.ls()` can be used to estimate separate ARX models. They all use `regressor.matrix()` which constructs, as the name suggests, a regressor matrix consisting of the lags of the dependent variable and (if specified) the exogenous variables. We once again want to stress that the model selection procedure assumes that the exogenous variables are chosen a priori. In the spirit of augmented distributed lag models, one might want to also include lags of the exogenous variables. Note that one can easily compute a regressor matrix of exogenous variables and their lags using the `regressor.matrix()` function. The exact input arguments of all functions can be found in the manual available at <https://CRAN.R-project.org/package=MARX>.

6.3.4 MARX models

As soon as the total autoregressive order p has been established, we want to select the MARX(r, s, q) model with $p = r + s$ that maximizes the log-likelihood at the estimated parameters. This information can immediately be attained using

```
> selection.lag.lead(y,x,3)
```

which produces the following output:

```
\$p.C
```

```
[1] 2
```

```
\$p.NC
```

```
[1] 1
```

```
\$loglikelihood
```

```
[1] -411.8389 -394.7271 -400.1387 -418.7731
```

Note that the output ‘loglikelihood’ is a list containing the log-likelihood of the models MARX(3,0,1), MARX(2,1,1), MARX(1,2,1) and MARX(0,3,1) respectively. The second element of this list has the highest log-likelihood which corresponds to $r = 2$ and $s = 1$. This is indeed in line with the model we used to simulate the data from. Similar to the pseudo-causal case, we now create an object of the **S3**-class “mixed” in order to easily access for example the estimated coefficients, residuals and fitted values:

```
final.model <- mixed(y,x,2,1)
```

Using the function `summary()` on this object, we obtain:

```
\$call
mixed.default(y = y, x = x, p_C = 2, p_NC = 1)

\$coefficients
      Estimate   Std.Err   t value   p value
int    -0.3392083 0.11042714 -3.071784 2.436788e-03
lag 1    0.2490939 0.03711930  6.710629 2.121500e-10
lag 2    0.3623430 0.03711883  9.761701 1.559521e-18
lead 1   0.5402226 0.03462631 15.601507 5.587446e-36
exo 1    0.5080777 0.03040126 16.712389 2.702860e-39

attr("class")
[1] "summary.mixed"
```

These estimated parameters are very close to the original coefficients chosen to simulate the original MARX model. Note that these are the estimates corresponding to estimating the model as defined in (6.1) augmented with an intercept. That is, **lag 1** and **lag 2** correspond to the estimates of the coefficients ϕ_1 and ϕ_2 of the polynomial $\phi(L)$, whereas **lead 1** is the estimate of φ_1 in $\varphi(L^{-1})$. Accordingly, the estimate **exo 1** is the estimated value for β . In finite samples, the simulation procedure is known to introduce an intercept when the error distribution exhibits fat tails. The functions `inference()` is used to compute the standard errors as proposed in Hecq et al. (2017a) and can also be used to access corresponding confidence intervals. The function `marx.t()` can be seen as the `arx.ls()` counterpart, only this function (i) estimates mixed causal-noncausal models instead of pseudo-causal models and (ii) uses ML estimation instead of OLS. It makes use of `ll.max()` which is the Student's t log-likelihood function that is maximized using `optim()` from the `stats` package. By filling in the right arguments, all these functions can also be used separately.

6.4 An Application on Commodity Prices

As has been demonstrated in the previous section, the **MARX** package is a set of integrated functions which makes the simulation, estimation and selection of mixed

causal-noncausal model very straightforward. However, we did not yet discuss the `marx()` function, which completely integrates the two steps of the model selection procedure. In this section, we will apply the function to commodity prices in relation to the U.S. exchange rate, similar to Hecq et al. (2017a). They consider the growth rate of commodity prices, as these series are believed to potentially contain a forward-looking component similar to more general measures of inflation. Moreover, as commodities are mainly priced in dollar, a negative relation between commodity prices and the U.S. exchange rate is expected. Let us read in the data that is supplied with the package:

```
> data <- MARX::dataset
```

The object “data” is a spreadsheet of 8 variables (columns) spanning from February 1980 until October 2016, which accounts for 441 observations (rows) for each variable. In this application we focus on the series “dlnoil” which is the monthly growth rate of the commodity crude oil and its relation to “dlnex” which is the monthly growth rate of the trade weighted U.S. dollar index, i.e. the U.S. exchange rate. These series are displayed in Figure 6.4.

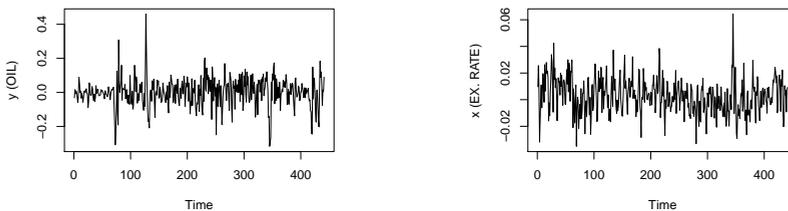


Figure 6.4: The monthly growth rate of commodity price oil (left) and U.S. dollar index (right)

The function `marx()` has 4 mandatory inputs: the variable y , the (matrix of) exogenous variable(s) x , the maximum number of autoregressive terms to be considered for model selection `p_max` and the significance level chosen for the computation of confidence intervals `sig_level`.¹¹ We choose `p_max = 6` and `sig_level = 0.05` and run the following code

```
> marx(data[,6],data[,8],8,0.05)
```

¹¹One can also specify two more inputs: `p_C` and `p_NC`, then the argument `p_max` becomes irrelevant and the function estimates and MARX with $r = p_C$ and $s = p_NC$.

which gives the following output:

```
marx(y = data[, 6], x = data[, 8], p_max = 6, sig_level = 0.05)
```

Order Selection Criteria Pseudo Causal Model:

BAYESIAN INFORMATION CRITERION

```
      p = 0    p = 1    p = 2    p = 3    p = 4    p = 5    p = 6
[1,] -2.20713 -2.28638 -2.27605 -2.25995 -2.25902 -2.24491 -2.24075
```

Minimum value attained at p = 1

AKAIKE INFORMATION CRITERION

```
      p = 0    p = 1    p = 2    p = 3    p = 4    p = 5    p = 6
[1,] -2.22568 -2.31425 -2.31326 -2.30655 -2.31503 -2.31038 -2.31570
```

Minimum value attained at p = 6

HANNAN-QUINN INFORMATION CRITERION

```
      p = 0    p = 1    p = 2    p = 3    p = 4    p = 5    p = 6
[1,] -2.21836 -2.30325 -2.29858 -2.28816 -2.29293 -2.28454 -2.28612
```

Minimum value attained at p = 1

Choose lag order for pseudo causal model:

This output reveals that the information criteria are not proposing the same lag order. Hence, we would like to stress that it is advisable to perform tests on the residuals of these pseudo-causal models to make sure the serial correlation has been captured adequately. Applying various tests and by looking at the correlogram, we choose $p = 4$ as total autoregressive order. Hence, we fill in **Choose lag order for pseudo causal model: 4**. After this has been specified, the program produces the following output:

```
THE KS-TEST REJECTS THE NULL OF NORMALLY DISTRIBUTED RESIDUALS
OF THE PURELY CAUSAL ARX MODEL p-value:0
```

```
THE JB-TEST REJECTS THE NULL OF NORMALLY DISTRIBUTED RESIDUALS
OF THE PURELY CAUSAL ARX MODEL p-value:0
```

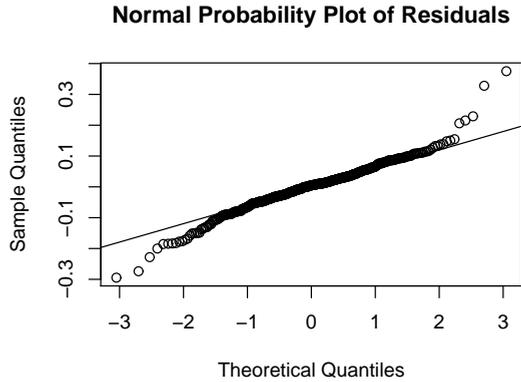


Figure 6.5: Normal Probability Plot of Residuals

Estimated Causal Parameters:

```
[1] -0.4866856 -0.1755051
```

Corresponding Standard Errors:

```
  [,1]
```

```
[1,] 0.03614241
```

```
[2,] 0.03617243
```

Corresponding Confidence Intervals:

```
  [,1]      [,2]
```

```
[1,] -0.5575234 -0.4158478
```

```
[2,] -0.2464018 -0.1046085
```

Estimated Noncausal Parameters:

```
[1] 0.7227104 -0.2372586
```

Corresponding Standard Errors:

```
  [,1]
```

```
[1,] 0.03489924
```

```
[2,] 0.03486819
```

Corresponding Confidence Intervals:

```
          [,1]      [,2]
[1,]  0.6543091  0.7911116
[2,] -0.3055990 -0.1689182
```

Estimated Parameters Exogenous Variables:

```
[1] -1.078717
```

Corresponding Standard Errors:

```
          [,1]
[1,]  0.2101618
```

Corresponding Confidence Intervals:

```
          [,1]      [,2]
[1,] -1.490626 -0.6668072
```

Estimated Intercept:

```
[1] 0.007987817
```

Corresponding Standard Errors:

```
[1] 0.002853031
```

Corresponding Confidence Intervals:

```
[1] 0.002395979 0.013579655
```

Estimated Distributional Parameters (df,sig):

```
[1] 2.88345912 -0.04845518
```

Figure 6.5 shows a Q-Q plot between the theoretical quantiles of the normal distribution and the sample quantiles from the pseudo-causal residuals. The sample quantiles deviate quite a bit from the theoretical quantiles of the normal, which indicates that it is likely that the residuals do not follow a normal distribution. In order to have more formal evidence, the function also outputs the results of the Kolmogorov-Smirnov test (`ks.test()` from the `stats` package) and the Jarque-Bera test (`jarque.bera.test()` from the `tseries` package) that test for the null of

normality. In both cases, the null hypothesis is rejected with a p -value almost completely equal to 0. The remaining output shows that the procedure chose the **MARX**(2,2,1) to be the model with the highest log-likelihood among all models with $p = r + s$. For all model parameters, it outputs the estimates, the corresponding standard errors and 95% confidence intervals (as we chose `sig.level = 0.05`). Finally, also the estimated distributional parameters are reported which indicate that the distribution is indeed more leptokurtic than the normal distribution ($\hat{\nu} = 2.88$) and rescaled ($\hat{\sigma} = -0.048$).

6.5 Conclusion

This chapter introduced the **R** package **MARX** for simulating, estimating and selecting mixed causal-noncausal autoregressive models with possibly exogenous regressors. It allows practitioners in many scientific fields to perform applied research using **MARX** models in a comprehensive, user-friendly environment.

We extensively illustrated the package usage and emphasized the ease with which it can be applied to real-life problems. We performed analysis on simulated data and a real empirical application using data on commodity prices and the U.S. exchange rate. The flexibility of the package allows the researcher to extract any relevant information (e.g., residuals, coefficients) easily from the classes “pseudo” and “mixed”, which makes it possible to analyze them further using other packages. If you use **R** or **MARX**, please cite the software in your publications.

Computational details

The results in this chapter were obtained using **R** 3.1.3 (R Core Team, 2017) with the packages: **matlab** version 1.0.2 (Roebuck, 2014), **fBasics** version 3011.87 (Rmetrics Core Team et al., 2014), **tseries** version 0.10-41 (Trapletti et al., 2017) and **stabledist** version 0.7-1 (Wuertz et al., 2016). These packages are specified as imports of the **MARX** package, which means that they are automatically installed together with **MARX**. All computations were performed on a Intel(R) Core(TM) i5-4570 CPU @ 3.20 GHz processor.

R itself and all packages used are available from **CRAN** at <http://CRAN.R-project.org/>. The package **MARX** is available from the **CRAN** repository at <https://cran.r-project.org/package=MARX>. The version under development is available in **GitHub** at <https://github.com/seantelg/MARX>.

Chapter 7

Conclusion

“All good things must come to an end.”

-Geoffrey Chaucer (1343-1400)

This chapter provides an overall conclusion of this thesis, more detailed conclusions can be found at the end of each individual chapter. Noncausal models are a well-known workhorse in various scientific fields (like e.g. physics and astronomy), but they have only recently received attention in the econom(etr)ic literature. The noncausal stationary solution of an autoregressive model is often considered “unnatural”, as it is difficult to find a clear intuition behind economic processes that solely depend on future values. However, from a statistical point of view, there is no reason to exclude these specifications from the set of potential models. This thesis shows how the introduction of noncausality can be useful in explaining various economic processes and phenomena. In particular, several macroeconomic and financial data are found to be represented more accurate (in terms of fit) by means of models that consist of a forward-looking component. New tools and extensions are proposed to accommodate better ways of modelling and analyzing mixed causal-noncausal models, for which routines are implemented in `matlab` and `R`. This new class of models is investigated in the context of important economic concepts like seasonal adjustment and the detection of common features.

Chapter 2 compares the finite sample behavior of the maximum likelihood estimator (assuming a t -distribution) and the least absolute deviation estimator. By means of simulations, it is found that both methods capture the causal and/or non-causal dependency particularly well when the distribution has very fat tails. Both estimators converge faster for infinite variance mixed causal-noncausal autoregressive processes and the identification is more accurate as well. The asymptotic distribution of the causal and noncausal autoregressive parameters is specified for t -distributed and Laplacian distributed error terms. This offers the possibility to compute standard errors in closed form, which is particularly useful as gradient-based numerical methods are found to be unstable in certain instances. This chapter also serves as a guide for practitioners to implement mixed causal-noncausal models step-by-step. These routines are implemented in `matlab` and `R`.

In **Chapter 3** we focus on seasonal adjustment techniques that might spuriously introduce a causal and/or noncausal component in time series. It is found that the X-11 seasonal adjustment filter, which is the basis of various widely used methods, directly affects the model selection of mixed causal-noncausal models. In all cases, the adjustment method introduces autocorrelations which leads to an overselection of models with higher-order dynamics. For a white noise series, the X-11 seasonal adjustment introduces both spurious causal and noncausal components with equal probability. For purely causal, noncausal and mixed causal-noncausal

processes, higher order models are selected and (non)causality is preserved in most cases. These simulation results are illustrated in an empirical application on inflation rates for the G7 countries. It is explained how discrepancies might arise due to the presence of ARCH effects and the potential flaws in the model selection procedure due to non-nestedness of mixed causal-noncausal processes.

Chapter 4 extends the mixed causal-noncausal autoregressive model with exogenous regressors and shows the consistency of the approximate non-Gaussian maximum likelihood estimator in this setup. This allows researchers to consider the direct effect of exogenous variables in structural economic relationships that are known to have explosive roots when represented by a causal autoregression. Whereas noncausal models are known to take into account the effect of omitted variables, the statistical significance of the coefficients of exogenous variables suggests that not all effect is taken into account. This is due to the fact that a noncausal model only includes the effect of exogenous variables that is predictable by its dependent variable. The introduction of exogenous regressors makes it possible to identify mixed causal-noncausal autoregressions from a second-order point of view due to the fact that cross-covariances are generally not symmetric.

In **Chapter 5** a novel explanation is given for the inability of detecting common features in a vector autoregressive framework. It is shown that common features that cannot be detected in the causal representation might be revealed in the noncausal dynamics of the series. Tools are proposed to identify the potential presence of noncausal co-movements. Both in a Monte Carlo study and two empirical applications additional relationships are discovered when noncausality is allowed. The presence of a reduced rank structure in the dynamics of purely causal and purely noncausal systems with more than one lag/lead ensures identification in the usual Gaussian framework. This result cannot be extended to mixed causal-noncausal specifications as the instruments are correlated with the error term of the reduced rank model by construction.

Chapter 6 presents the **MARX** package for the analysis of mixed causal-noncausal autoregressions with exogenous variables. The software package offers functions to simulate, estimate and select mixed causal-noncausal models along the lines of Chapter 4. It allows practitioners in many scientific fields to perform applied research on this set of models in a comprehensive, user-friendly environment.

To conclude, the mixed causal-noncausal model has practical relevance in econom(etr)ics because of various reasons. They are relatively simple models that

are able to generate features that previously could only be obtained using highly nonlinear and complex models. This allows for accurate modelling of processes, especially those that contain asymmetric cycles and speculative bubbles. The introduction of exogenous variables in the mixed causal-noncausal model offers even more flexibility, as one can include features and shocks that cannot be explained by past and future values of the dependent variable. As these models are only recently introduced in the literature, there are still various settings and contexts in which their behavior and implications can be studied.

Bibliography

*“Research is to see what everybody else has seen,
and to think what nobody else has thought.”*

-Albert Szent-Gyorgyi (1893-1986)

-
- [1] ABELN, B. AND J. JACOBS (2015), Seasonal Adjustment With and Without Revisions: A Comparison of X-13ARIMA-SEATS and CAMPLET, *CAMA Working Papers 2015-25*.
- [2] ALESSI, L., BARIGOZZI, M. AND M. CAPASSO (2011), Non-Fundamentalness in Structural Econometric Models: A Review, *International Statistical Review*, 79(1), 16-47.
- [3] ANDERSEN, T.G., DOBREV, D. AND E. SCHAUMBURG (2010), Jump-Robust Volatility Estimation Using Nearest Neighbor Truncation, *Staff Report*, 465, Federal Reserve Bank of New York.
- [4] ANDREWS, B., DAVIS, R. AND F. BREIDT (2006), Maximum Likelihood Estimation For All-Pass Time Series Models, *Journal of Multivariate Analysis*, 97, 1638-1659.
- [5] ANDREWS, B. AND R. DAVIS (2013), Model Identification for Infinite Variance Autoregressive Processes, *Journal of Econometrics*, 172, 222-234.
- [6] BARNDORFF-NIELSEN, O. E. AND N. SHEPHARD (2004), Power and Bipower Variation with Stochastic Volatility and Jumps (with Discussion), *Journal of Financial Econometrics*, 2, 1-37.
- [7] BAUWENS, L., HAFNER, CH. AND S. LAURENT (2012), Volatility Models, *Handbook of Volatility Models and Their Applications*, Chapter 1, Wiley.
- [8] BEAUDRY, P., FÈVE, P., GUAY, A. AND F. PORTIER (2015), When is Nonfundamentalness in VARs a Real Problem? An Application to News Shocks, *CEPR Discussion Paper*, 10763.
- [9] BELL, W. (1987), A Note on Overdifferencing and the Equivalence of Seasonal Time Series Models with Monthly Means and Models with $(0, 1, 1)_{12}$ Seasonal Parts when $\Theta = 1$, *Journal of Business and Economic Statistics*, 5, 383-387.
- [10] BELL, W. AND S. HILLMER (2002), Issues Involved with the Seasonal Adjustment of Economic Time Series, *Journal of Business & Economic Statistics*, 20(1), Twentieth Anniversary Commemorative Issue, 98-127.
- [11] BORK, L., KALTWASSER, P. AND P. SERCU (2014), Do Exchange Rates Really Help Forecasting Commodity Prices?, *Working Paper*, available at SSRN: <http://ssrn.com/abstract=2473624>.

- [12] BREIDT, F. AND R. DAVIS (1992), Time-Reversibility, Identifiability and Independence of Innovations for Stationary Time Series, *Journal of Time Series Analysis*, 13(5), 377-390.
- [13] BREIDT, F., DAVIS, R., LIU, K. AND M. ROSENBLATT (1991), Maximum Likelihood Estimation for Noncausal Autoregressive Processes. *Journal of Multivariate Analysis*, 36, 175-198.
- [14] BROCKWELL, P. AND R. DAVIS (1991), *Time Series: Theory and Methods*, Springer-Verlag New York, Second Edition.
- [15] BROCKWELL, P. AND R. DAVIS (2002), *Introduction to Time Series and Forecasting*, Springer-Verlag New York, Second Edition.
- [16] BROZE, L., GOURIÉROUX, C. AND A. SZAFARZ (1995), Solutions of Multivariate Rational Expectation Models, *Econometric Theory*, 11, 229-257.
- [17] CANDELON, B., HECQ, A. AND W. VERSCHOOR (2005), Measuring Common Cyclical Features During Financial Turmoil, *Journal of International Money and Finance*, 24, 1317-1334
- [18] CASELLA, G. AND R. BERGER (2002), *Statistical Inference*, Thomson Learning, Second Edition.
- [19] CAVALIERE, G., NIELSEN, H.B. AND A. RAHBEK (2017), Bootstrapping Non-Causal Autoregressions: with an Application to Explosive Bubble Modelling. *Working Paper*, University of Copenhagen, Denmark.
- [20] CHEN, Y., ROGOFF, K. AND B. ROSSI (2010), Can Exchange Rates Forecast Commodity Prices?, *The Quarterly Journal of Economics*, 125(3), 1145-1194.
- [21] CORRADI, V. AND N. SWANSON (2006), Predictive Density Evaluation, *Handbook of Economic Forecasting*, Chapter 5, Vol. 1, 197-284.
- [22] CORSI, F. (2009), A Simple Approximate Long-Memory Model of Realized Volatility, *Journal of Financial Econometrics*, Vol. 7, No.2, 174-196.
- [23] CORSI, F., AUDRINO, F. AND R. RENÒ (2012), HAR Modeling for Realized Volatility Forecasting, *Handbook of Volatility Models and Their Applications*, Wiley, Ch. 15, Ed. by Bauwens, Hafner and Laurent.
- [24] CUBADDA, G. (1999), Common Cycles in Seasonal Non-Stationary Time Series, *Journal of Applied Econometrics*, 14, 273-291.

-
- [25] CUBADDA, G. (2007), A Unifying Framework for Analysing Common Cyclical Features in Cointegrated Time Series, *Computational Statistics & Data Analysis*, 52, 896–906.
- [26] CUBADDA, G., AND A. HECQ (2001), On Con-Contemporaneous Short-Run Co-Movements, *Economics Letters*, 73, 389-397.
- [27] CUBADDA, G., HECQ, A. AND S. TELG (2017), Serial Correlation Common Noncausal Features, *MPRA Paper 77254*, University Library of Munich, Germany.
- [28] DAVIS, R., KNIGHT, K. AND J. LIU (1992), M-Estimation for Autoregressions with Infinite Variance, *Stochastic Processes and Their Applications*, 40, 145-180
- [29] DAVIS, R. AND L. SONG (2012), Noncausal Vector AR Processes with Application to Economic Time Series, *Discussion Paper Columbia University*.
- [30] DE JONG, R. AND N. SAKARYA (2016), The Econometrics of the Hodrick-Prescott Filter, *The Review of Economics and Statistics*, 98(2), 310-317.
- [31] DEL BARRIO CASTRO, T. AND D. OSBORN (2004), The Consequences of Seasonal Adjustment for Periodic Autoregressive Processes, *Econometrics Journal*, 7, 207-231.
- [32] DEL BARRIO CASTRO, T., RODRIGUES P. AND R. TAYLOR (2017), Semi-Parametric Seasonal Unit Root Tests, *Econometric Theory*, 1-30, doi:10.1017/S0266466617000135.
- [33] ENGLE, R. F. AND S. KOZICKI (1993), Testing for Common Features (with Comments), *Journal of Business and Economic Statistics* 11, 369-395.
- [34] FONSECA, T., FERREIRA, M. AND H. MIGON (2008), Objective Bayesian Analysis for the Student-*t* Regression Model, *Biometrika*, 95, 325-333.
- [35] FRANCO, C. AND J.-M. ZAKOÏAN (2007), HAC Estimation and Strong Linearity Testing in Weak ARMA Models, *Journal of Multivariate Analysis* 98, 114-144.
- [36] FRANSSES, PH. H. AND B. HOBIJN (1997), Critical Values for Unit Root Tests in Seasonal Time Series, *Journal of Applied Statistics*, 24, 25-47.

- [37] FRIES, S. AND J.-M. ZAKOÏAN (2017), Mixed Causal-Noncausal AR Processes and the Modelling of Explosive Bubbles. *MPRA Paper 81345*, University Library of Munich, Germany.
- [38] GHYSELS (1990), Unit Root Tests and the Statistical Pitfalls of Seasonal Adjustment: The Case of U.S. Post War Real GNP, *Journal of Business and Economic Statistics*, 8, 145-152.
- [39] GHYSELS, E., LEE H. AND P. SIKLOS (1993), On the (Mis)specification of Seasonality and its Consequences: An Empirical Investigation with U.S. Data, *Empirical Economics*, 18(4), 191-204.
- [40] GHYSELS, E. AND P. PERRON (1993), The Effect of Seasonal Adjustment Filters on Tests for a Unit Root, *Journal of Econometrics*, 55, 57-98.
- [41] GOURIÉROUX, C. AND J. JASIAK (2016), Filtering, Prediction and Simulation Methods in Noncausal Processes. *Journal of Time Series Analysis*, 37, 405-430. doi: 10111/jtsa.12165.
- [42] GOURIÉROUX, C. AND J. JASIAK (2017A), Misspecification of Causal and Noncausal Orders in Autoregressive Processes, *Journal of Econometrics*, forthcoming.
- [43] GOURIÉROUX, C. AND J. JASIAK (2017B), Noncausal Vector Autoregressive Processes: Representation, Identification and Semi-Parametric Estimation, *Journal of Econometrics*, forthcoming.
- [44] GOURIÉROUX, C., JASIAK, J. AND A. MONFORT (2016), Stationary Bubble Equilibria in Rational Expectation Models, *CREST Working Paper*, 2016-31.
- [45] GOURIÉROUX, C., AND J.M. ZAKOÏAN (2013), Explosive Bubble Modeling by Noncausal Process, *CREST Discussion Papers 2013-04.*, Version updated in August, 2014.
- [46] GOURIÉROUX, C. AND J.M. ZAKOÏAN (2014), On Uniqueness of Moving Average Representations of Heavy-Tailed Stationary Processes, *Journal of Time Series Analysis*, 36, 876-887.
- [47] GOURIÉROUX, C., AND J.M. ZAKOÏAN (2016), Local Explosion Modelling by Noncausal Process, *Journal of the Royal Statistical Society, Series B*, doi:10.1111/rssb.12193.

-
- [48] GUILLÉN, O., HECQ, A., ISSLER, J. AND D. SARAIVA (2015), Forecasting Multivariate Time Series Under Present-Value Model Short- and Long-Run Co-Movement Restrictions, *International Journal of Forecasting, Elsevier*, 31(3), 862-875.
- [49] HAMILTON, J. (1994), *Time Series Analysis*, Princeton University Press.
- [50] HANNAN, E., DUNSMUIR W. AND M. DEISTLER (1980), Estimation of Vector ARMAX Models, *Journal of Multivariate Analysis*, 10(3), 275-295.
- [51] HANSEN, L. (1982), Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, 50(4), 1029-1054.
- [52] HANSEN, L. AND T. SARGENT (1991), Two Difficulties in Interpreting Vector Autoregressions, *Rational Expectation Econometrics*, 77-119, Westview Press, Boulder.
- [53] HEBER, G., LUNDE, A., SHEPHARD, N. AND K. SHEPPARD (2009), Oxford-Man Institute's Realized Library, Version 0.1, *Oxford-Man Institute, University of Oxford*.
- [54] HECQ, A. (1996), IGARCH Effects on Autoregressive Lag Length Selection and Causality Tests, *Applied Economic Letters*, 3, 317-323.
- [55] HECQ, A. (1998), Does Seasonal Adjustment Induce Common Cycles?, *Economics Letters*, 59, 289-297.
- [56] HECQ, A. AND J.V. ISSLER (2012), A Common-Feature Approach for Testing Present-Value Restrictions with Financial Data, *METEOR RM/12/006*, Maastricht University.
- [57] HECQ, A., ISSLER, J.V. AND S. TELG (2017A), Mixed Causal-Noncausal Autoregressive Processes with Strictly Exogenous Regressors, *MPRA Paper 80767*, University Library of Munich, Germany.
- [58] HECQ, A., LIEB, L. AND S. TELG (2016), Identification of Mixed Causal-Noncausal Models in Finite Samples, *Annals of Economics and Statistics*, 123/124, 307-331.
- [59] HECQ, A., LIEB, L. AND S. TELG (2017B), Simulation, Estimation and Selection of Mixed Causal-Noncausal Autoregressive Models: The **MARX**

- Package, *Working Paper*, available at SSRN: <https://ssrn.com/abstract=3015797>.
- [60] HECQ, A., PALM F.C. AND J.P. URBAIN (2006), Testing for Common Cyclical Features in VAR Models with Cointegration, *Journal of Econometrics* 132(1), 117-141.
- [61] HECQ, A., TELG, S. AND L. LIEB (2017C), Do Seasonal Adjustments Induce Noncausal Dynamics in Inflation Rates?, *Econometrics*, 5(4), 48, doi:10.3390/econometrics5040048.
- [62] HENCIC, A. AND C. GOURIÉROUX (2014), Noncausal Autoregressive Model in Application to Bitcoin/USD Exchange Rate, *Econometrics of Risk, Series: Studies in Computational Intelligence*, Springer International Publishing, 17-40.
- [63] HOMM, U. AND J. BREITUNG (2012), Testing for Speculative Bubbles in Stock Markets: a Comparison of Alternative Methods, *Journal of Financial Econometrics*, 10(1), 198-231.
- [64] HUANG J. AND Y. PAWITAN (2000), Quasi-likelihood Estimation of Non-invertible Moving Average Processes, *Scandinavian Journal of Statistics*, 27, 689-710.
- [65] HURVICH, M. AND C.L. TSAI (1989), Regression and Time Series Model Selection in Small Samples, *Biometrika*, 76, 297-307.
- [66] HYLLEBERG, S., ENGLE, R., GRANGER C. AND B. YOO (1990), Seasonal Integration and Cointegration, *Journal of Econometrics*, 44, 215-238.
- [67] ISSLER, J. AND F. VAHID (2001), Common Cycles and the Importance of Transitory Shocks to Macroeconomic Aggregates, *Journal of Monetary Economics*, 47(3), 449-475.
- [68] JACKSON, L.B. (1989), Noncausal ARMA Modeling of Voiced Speech, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(10), 1606-1608.
- [69] JOHANSEN, S. (2008), Reduced Rank Regression, *The New Palgrave Dictionary of Economics* (2nd Edition), Durlauf, S.N. and L.E. Blume (eds.), doi:10.1057/9780230226203.1410.

-
- [70] KAISER, R. AND A. MARAVALL (2001), *Measuring Business Cycles in Economic Time Series*, Springer, New York.
- [71] KNIGHT K. (1998), Limiting Distributions for L_1 Regression Estimators under General Conditions, *The Annals of Statistics*, 26(2), 755-770.
- [72] LANNE, M. AND J. LUOTO (2012), Has U.S. Inflation Really Become Harder To Forecast?, *Economics Letters*, 155(3), 383-386.
- [73] LANNE, M. AND J. LUOTO (2013), Autoregression-Based Estimation of the New Keynesian Phillips Curve, *Journal of Economic Dynamics & Control*, 37, 561-570.
- [74] LANNE, M., LUOTO J. AND P. SAIKKONEN (2012A), Optimal Forecasting of Noncausal Autoregressive Time Series, *International Journal of Forecasting*, 28, 623-631.
- [75] LANNE, M., NYBERG, H. AND E. SAARINEN (2012B). Does Noncausality Help in Forecasting Economic Time Series?, *Economics Bulletin*, 32(4), 2849-2859.
- [76] LANNE, M. AND P. SAIKKONEN (2011A), Noncausal Autoregressions for Economic Time Series, *Journal of Time Series Econometrics*, 3(3), 1-32.
- [77] LANNE, M. AND P. SAIKKONEN (2011B), GMM Estimation with Noncausal Instruments, *Oxford Bulletin of Economics and Statistics*, 73(5), 581-592.
- [78] LANNE, M. AND P. SAIKKONEN (2013), Noncausal Vector Autoregression, *Econometric Theory*, 29(3), 447-481.
- [79] LIU, K. AND ROSENBLATT, M. (1993), Non-Gaussian Autoregressive Moving Average Processes, *Proceedings of the National Academy of Science of the United States of America*, 90(19), 9168-9170.
- [80] LIPPI, M. AND L. REICHLIN (1993A), The Dynamic Effects of Aggregate Demand and Supply Disturbances, *American Economic Review*, 83(3), 644-52.
- [81] LIPPI, M. AND L. REICHLIN (1993B), VAR Analysis, Nonfundamental Representations, Blaschke Matrices, *Journal of Econometrics*, 63, 307-325.
- [82] LOF, M. AND H. NYBERG (2017), Noncausality and the Commodity Currency Hypothesis, *Energy Economics*, 65, 424-433.

- [83] MARAVALL, A. (1993), Stochastic Linear Trends, Models and Estimators, *Journal of Econometrics*, 56, 5-37.
- [84] MARAVALL, A. (1997), Use and (Mostly) Abuse of Time Series Techniques in Economic Analysis, *Paper Presented at the 5th CEMAPRE Conference, Lisboa*.
- [85] MARAVALL, A. (2006), An Application of the TRAMO-SEATS Automatic Procedure; Direct Versus Indirect Adjustment, *Computational Statistics & Data Analysis*, 50(9), 2167-2190.
- [86] MCCracken, M.W. AND S. NG (2016), FRED-MD: A Monthly Database for Macroeconomic Research, *Journal of Business & Economic Statistics*, 34(4), 574-589, doi:10.1080/07350015.2015.1086655.
- [87] NELSON, C. AND H. KANG (1981), Spurious Periodicity in Inappropriately Detrended Time Series, *Econometrica*, 49(3), 741-751.
- [88] PESARAN, H. (2015), *Time Series and Panel Data Econometrics*, Oxford University Press.
- [89] PHILLIPS, P.C., WU, Y. AND J. YU (2011), Explosive behavior in the 1990s Nasdaq: When Did Exuberance Escalate Asset Values?, *International Economic Reviews*, 52(1), 201-226.
- [90] RAMSEY, J. AND P. ROTHMAN (1996), Time Irreversibility and Business Cycle Asymmetry, *Journal of Money, Credit and Banking*, 28(1), 1-21.
- [91] R CORE TEAM (2017), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>.
- [92] REINSEL, G. (1997), *Elements of Multivariate Time Series Analysis*, Springer Science+Business Media New York, Second Edition.
- [93] RMETRICS CORE TEAM, WUERTZ, D., SETZ, T. AND Y. CHALABI (2014), **fBasics**: *Rmetrics - Markets and Basic Statistics*, R package version 3011.87, <https://CRAN.R-project.org/package=fBasics>.
- [94] ROEBUCK, P. (2014), **matlab**: *MATLAB emulation package*, R package version 1.0.2, <https://CRAN.R-project.org/package=matlab>.

- [95] ROSENBLATT, M. (2000), *Gaussian and Non-Gaussian Linear Time Series and Random Fields*, Springer-Verlag New York.
- [96] SAIKKONEN, P. AND R. SANDBERG (2016), Testing for a Unit Root in Non-causal Autoregressive Models, *Journal of Time Series Analysis*, 37(1), 99-125.
- [97] SCARGLE, J.D. (1981), Studies in Astronomical Time Series Analysis. I-Modeling Random Processes in the Time Domain”, *Astrophysical Journal Supplement Series*, 45, 1-71.
- [98] SCHMIDT, H. (1978), Can an Effect Precede Its Cause? A Model of a Non-causal World, *Foundations of Physics*, 8(5-6), 463-480.
- [99] SILVERMAN, B. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London – New York, 175
- [100] SCHERBINA, A. (2013), Asset Price Bubbles: A Selective Survey, *IMF working paper WP/13/45*.
- [101] STIGLITZ, J. (1990), Symposium on Bubbles, *Journal of Economic Perspectives*, Vol. 4, No. 2, 13-18.
- [102] TAKEUCHI, K. (1976), Distribution of informational statistics and a criterion of model fitting, *Suri-Kagaku (Mathematic Sciences)*, 153, 12-18 (in Japanese).
- [103] TRAPLETTI, A., HURNIK, K. AND B. LEBARON (2017), **tseries**: *Time Series Analysis and Computational Finance*, R package version 0.10-41, <https://CRAN.R-project.org/package=tseries>.
- [104] VAHID, F. AND R. ENGLE (1993), Common Trends and Common Cycles, *Journal of Applied Econometrics*, 8(4), 341-360.
- [105] VAHID, F. AND R. ENGLE (1997), Codependent cycles, *Journal of Econometrics*, 80(2), 199-221.
- [106] WEI, W. (2006), *Time Series Analysis*, Pearson Addison Wesley.
- [107] WU R. AND R. DAVIS (2010), Least Absolute Deviation Estimation for General Autoregressive Moving Average Time-Series Models. *Journal of Time Series Analysis*, 31, 98-112.

- [108] WUERTZ, D., MAECHLER M., AND RMETRICS CORE TEAM (2016), **stabledist**: *Stable Distribution Functions*, **R** package version 0.7-1, <https://CRAN.R-project.org/package=stabledist>.
- [109] YANAGIHARA, H. (2006), Corrected version of AIC for selecting multivariate normal linear regression models in a general nonnormal case, *Journal of Multivariate Analysis*, 97, 1070-1089.

Nederlandse Samenvatting

“You can never understand one language until you understand at least two.”

-Geoffrey Willans (1911-1958)

Dit proefschrift bestudeert het causale/non-causale autoregressief model, aangeduid als MAR, en de economische toepassingen waarvoor het kan worden gebruikt. De introductie van een vooruitziend element leidt tot het abandonneren van de gebruikelijke Box-Jenkins methodiek om tijdreeksen te modelleren. Vooral de veronderstelling dat de storingsterm normaal verdeeld is, dient te worden opgegeven om identificatie te bewerkstelligen. Dit heeft tot direct gevolg dat de bestaande representatie theorie alsmede schattings- en testmethoden heroverwogen en aangepast dienen te worden voor causale/non-causale modellen. Deze dissertatie heeft twee doelen. Het eerste doel behelst het toegankelijk maken van deze modellen voor empirische onderzoekers. Dit wordt gerealiseerd door het beschikbaar stellen van ontwikkelde software om eenvoudig deze causale/non-causale modellen te schatten en te selecteren. Tevens wordt de bruikbaarheid getoetst in verscheidene Monte Carlo simulaties en empirische toepassingen op echte macro-economische en financiële data. Het tweede doel bestaat uit het uitbreiden van reeds bestaande modellen, schatters, testen en andere methoden met betrekking tot gevolgtrekking.

Het eerste hoofdstuk bevat een uitgebreide literatuurstudie, waarbij de voornaamste kenmerken en toepassingen van het MAR model worden gepresenteerd. In het bijzonder wordt stilgestaan bij het feit dat een explosief autoregressief proces in kalendertijd kan worden geïnterpreteerd als een stationair proces in omgekeerde tijd, daar de unieke stationaire oplossing afhangt van huidige en toekomstige storingstermen. Hiertoe worden economische variabelen die afhankelijk zijn van aanstaande verwachtingen (zoals inflatie) vaak beter gemodelleerd door de toevoeging van een non-causaal element. Het MAR model introduceert een set van non-fundamentele oplossingen voor structurele economische modellen. De introductie van deze toekomstige informatie zorgt in vele empirische applicaties voor betere voorspellingen van data. Dit kan ook verklaard worden door het feit dat non-causale modellen, in tegenstelling tot puur causale modellen, complexe dynamische patronen (bijvoorbeeld speculatieve bubbels en asymmetrische cycli) in tijdreeksen kunnen nabootsen. Voorgenoemde thematiek wordt uitvoerig besproken en geïllustreerd.

Hoofdstuk twee kan worden gezien als een gids voor empirische onderzoekers om succesvol MAR modellen te implementeren. Daartoe worden er software programma's aangeboden die dit soort modellen schatten en selecteren volgens het voorgestelde stappenplan. De werking van dit plan wordt geïllustreerd in twee empirische applicaties, bestaande uit het modelleren van een zonnepanelen-bubbel in Wallonië (België) en financiële aandelenindices van verschillende landen. De wer-

king van twee schattingsmethoden, de t -Maximum-Likelihood Estimator (t -MLE) en de Least Absolute Deviation (LAD) schatter, wordt onderzocht in situaties waar essentiële veronderstellingen, die ten grondslag liggen aan het model, worden geschonden. Voor beide schatters wordt tevens de asymptotische distributie afgeleid die een gesloten analytische vorm bezit, waardoor instabiele numerieke methoden niet langer noodzakelijk zijn om (bijvoorbeeld) standaardfouten te berekenen.

In het derde hoofdstuk wordt de aandacht verlegd naar seizoencorrectie van data en onderzocht hoe standaard methoden (gebaseerd op lineaire filters) de selectie van MAR modellen beïnvloeden. Resultaten in simulatie studies en een empirische applicatie op inflatie data van de G7 landen tonen aan dat de onderzochte methoden voor seizoencorrectie nieuwe dynamische patronen introduceren in tijdreeksen. Deze patronen zijn zowel causaal als non-causaal en leiden tot de selectie van dynamischere modellen. De voorspellingen van deze modellen zijn vaak zeer verschillend van de geselecteerde modellen voor ruwe data. Foutieve seizoencorrectie kan hierdoor grote gevolgen hebben voor economisch beleid.

In hoofdstuk vier wordt het MAR model uitgebreid met exogene regressoren (MARX). Deze uitbreiding maakt het mogelijk om standaard schattingsmethoden, zoals ordinary least squares (OLS), ook toe te passen op causale/non-causale modellen, daar de kruiscovarianties identificatie bewerkstelligen. Er wordt aangetoond dat het MARX model kan worden gezien als een nieuwe set oplossingen voor rationele verwachtingsmodellen. Een groot voordeel is dat de impact van exogene regressoren op de afhankelijke variabele direct waarneembaar is. De consistentie van de maximum-likelihood schatter wordt aangetoond en een asymptotische distributie afgeleid voor de t -MLE en LAD schatter, die beide een gesloten analytische vorm bezitten. De compatibiliteit van het stappenplan van hoofdstuk twee wordt getest voor het MARX model in Monte Carlo simulaties en geïllustreerd in een empirische applicatie die de relatie tussen grondstofprijzen en de wisselkoers beschouwt. Tevens wordt ook voor het modelleren van MARX processen een software programma aangeboden.

Hoofdstuk vijf beschouwt het meerdimensionale causale/non-causale model, aangeduid als MVAR, en biedt een nieuwe uitleg voor het probleem dat conventionele testen gemeenschappelijke kenmerken in tijdreeksen niet detecteren terwijl deze wel aanwezig zijn. Er wordt aangetoond dat de gemeenschappelijke kenmerken aanwezig kunnen zijn in het non-causale gedeelte van de dataset, welke in vele gevallen niet ontdekt kunnen worden door conventionele causale modellen. Bestaande testen worden uitgebreid om ook compatibel te zijn voor non-causale

tijdreeksen. De aanwezigheid van verminderde rang dynamiek in, hetzij het causale deel hetzij het non-causale deel, bewerkstelligt identificatie van puur causale en puur non-causale modellen met normaal-verdeelde storingstermen wanneer de dynamische order hoger is dan één.

In het laatste hoofdstuk wordt het **R** software pakket **MARX** besproken. Het kan gebruikt worden voor de analyse van MARX processen. Het software pakket bestaat uit verschillende functies die het mogelijk maken om causale/non-causale autoregressieve modellen (zowel met als zonder exogene variabelen) te simuleren, te schatten en te selecteren. De geïmplementeerde methoden zijn gebaseerd op de theorie van hoofdstuk vier. Het pakket is op een gebruiksvriendelijke wijze samengesteld zodat onderzoekers van verschillende wetenschappelijke disciplines relatief eenvoudig toegepast onderzoek kunnen doen met deze modellen.

Dit proefschrift hoopt een waardevolle bijdrage te leveren aan de literatuur over tijdreeksen met een vooruitziend element. Het causale/non-causale autoregressief model lijkt toepasbaar op vele macroeconomische en financiële data. Dit komt voornamelijk door het feit dat deze modellen dynamische patronen kunnen genereren die voorheen slechts met nonlineaire, complexe modellen konden worden gecreëerd. Hopelijk kunnen de resultaten in dit proefschrift zowel toegepaste als theoretische onderzoekers verder op weg helpen.

Valorization

“Strive not to be a success, but rather to be of value.”

-Albert Einstein (1879-1955)

Knowledge valorization refers to the utilization of scientific knowledge in practice. While reading a scientific work, one may ask questions like: “*what is the societal relevance?*” and “*what is its contribution to real life phenomena?*”. In this addendum, I outline the knowledge valorization of modelling time series processes by means of mixed causal-noncausal autoregressive models, which is the main topic of this dissertation. In particular, I discuss the social and economic relevance of the topic and the respective results, potential target groups, translated services, processes and activities, as well as the innovativeness and implementability of the research.

As discussed in the Introduction, noncausal autoregressive models are common in various scientific fields (e.g., physics and astronomy) but relatively novel in the field of econom(etr)ics. In this dissertation, the noncausal model as well as a combination of causal and noncausal models (i.e., the mixed causal-noncausal model) is applied to various macroeconomic and financial time series, often leading to a better in-sample fit than conventional autoregressive models. This is due to the fact that these models can capture certain nonlinear dynamics that previously could only be picked up by very complex models. In other words, mixed causal-noncausal models are able to replicate features such as speculative bubbles and asymmetric cycles that are often present in economic data. Especially over the last years, one can find various time series displaying patterns that resemble a bubble (e.g., dot-com bubble, American housing crisis, bitcoin and property bubbles in various countries) and thus it is fair to argue that there is a need for relatively simple models that can accommodate such features. To that end, theoretical and empirical research on these models can improve the understanding of certain economic and financial processes. From a very general point of view, anyone interested in explaining these types of data series can benefit from a thorough study on these models. If statistical agencies and policy makers decide to adopt the mixed causal-noncausal autoregressive model, it can be argued that its influence is extended to a much larger crowd, who benefits either directly or indirectly from the research in this field.

The answer to the question whether the mixed causal-noncausal model is going to be broadly adopted in the future, is not known at this point in time. Allowing for noncausality seems to be a rather natural way to extend the Box-Jenkins methodology, which has been one of the most common approaches in time series model selection since the seventies. It extends the existing procedure in the sense that the set of possible dynamic models that can be chosen increases greatly.

Hence, it offers practitioners more flexibility in modelling time series of interest. In case this extension of the Box-Jenkins approach becomes standard, research in this field can constitute to new topics taught to students in the field of econometric theory.

In the current literature, one can find many theoretical results on the mixed causal-noncausal autoregressive model. In particular, Rosenblatt (2000) provides an extensive investigation of stationary autoregressive and moving average linear sequences and random fields in the non-Gaussian context. This monograph has inspired researchers to introduce noncausality in the world of econ(etr)ics. Since non-Gaussianity of the process is required to distinguish between purely causal, noncausal and mixed causal-noncausal models, it has been an important aspect in the literature. More specifically, one can distinguish between two different streams: the Gaussianity assumption of the error distribution is abandoned and rather assumed to be (i) a nonstandardized Student's t -distribution and the model is estimated by maximum likelihood or (ii) an α -stable distribution (or the weaker assumption that the law of the errors belongs to the domain of attraction of a stable distribution) and the standard least-squares method is adhered to. Both approaches have their merits but also have clear disadvantages. Most notably, the first case requires a (potentially misspecified) parametric assumption on the error distribution. In the second case, the process is assumed to have (at least) infinite variance, which complicates the analysis (as it heavily affects asymptotic results) and might be an unrealistic assumption. This only illustrates that more theoretical research is necessary to make clever decisions in modelling mixed causal-noncausal processes.

This dissertation aims to offer a contribution to this theoretical study but mostly focuses on the empirical side of it. In contrast to the various papers that establish the theoretical foundations of the mixed causal-noncausal model, the number of applied papers is very limited. This dissertation fills this gap in the literature in three different ways. Firstly, the finite sample performance of different estimators is studied and closed-form solutions to compute standard errors are provided (Chapter 2 and 4). Whereas asymptotic results are very important in the field of time series econometrics, practitioners often experience different types of problems in empirical work. These chapters address those issues and provide a clear guidance on how to swiftly implement mixed causal-noncausal models (potentially including exogenous regressors). Secondly, the mixed causal-noncausal model is studied in relation to two very important econometric concepts: seasonal

adjustment and common features (Chapter 3 and 5). In the first chapter, the effect of seasonal adjustment methods on the selection of mixed causal-noncausal models is considered. Seasonal adjustment of time series is a tedious procedure that might introduce unwanted side-effects. This can lead to a wrong assessment of the importance of forward- and backward-looking behavior of variables of interest, which can have important consequences for e.g. forecasting and policy implementation. The second chapter investigates whether more common features between series can be discovered when one allows for noncausality. The detection of common features is important as it can unravel links between different variables, which is of interest for various reasons (e.g., policy measures, spillover effects, forecasting). Thirdly, routines are developed which make it easy for practitioners to simulate, estimate and select mixed causal-noncausal models (Chapter 6). More specifically, an **R** package is freely available online which attempts to reduce the gap between theoretical and applied results in this field considerably.

Despite their immediate interest for macroeconomic and financial researchers, the contributions presented in this thesis are not restricted to that particular field. The **R** package can be used in various studies which intend to fit mixed causal-noncausal models to their data of interest (e.g., physics and astronomy mentioned before). Moreover, as researchers only have a limited amount of data available, the finite sample results can also prove to be useful for practitioners in several fields. Even the results for the effect of seasonal adjustment on model selection of mixed causal-noncausal models can be generalized beyond the exercise of seasonal adjustment. The extensive simulation study of Chapter 3 shows how linear filters, which are often applied to data (e.g., taking average of three months to compute a quarterly observation), affect the choice for the final model. In addition, the tests proposed to detect common features do not specify a certain type of commonality. Whereas this dissertation focuses on common features in economic data, there is no reason to expect that the tests fail to identify any other types of common features.

Most importantly, it seems that mixed causal-noncausal models offer a relatively simple alternative to highly nonlinear and complex models. In general, it appears that researchers face the following trade-off when they want to model processes with complex dynamics: either one favors a complex (nonlinear) model with (simple) Gaussian disturbances or one chooses a simple (linear) model with complex (e.g. infinite variance) disturbances. So far, the focus in economics was mostly on the former option. The mixed causal-noncausal model now provides the

latter option and seems to be rather successful in modelling economic processes. However, there are still many avenues for future research. In particular, the non-nestedness of mixed causal-noncausal models introduces severe complications for model selection. Additionally, standard tests in time series econometrics (e.g., unit root tests) are (or might be) no longer applicable in the presence of noncausality and thus have to be re-examined, which hopefully leads to the development of an appropriate set of tools. In terms of the model, one could think of various interesting extensions, for example generalized autoregressive conditional heteroskedasticity (GARCH) and vector error correction (VECM) models. For additional proposed future research, I would like to refer the reader to the conclusions of the respective chapters. I hope that the methods considered in this dissertation are also employed outside of the academic world. As a last remark, I want to point out an important part of valorization that is often overlooked: writing this thesis has made me a better researcher and teacher, which - hopefully - also has a significant impact on society.

Curriculum Vitae

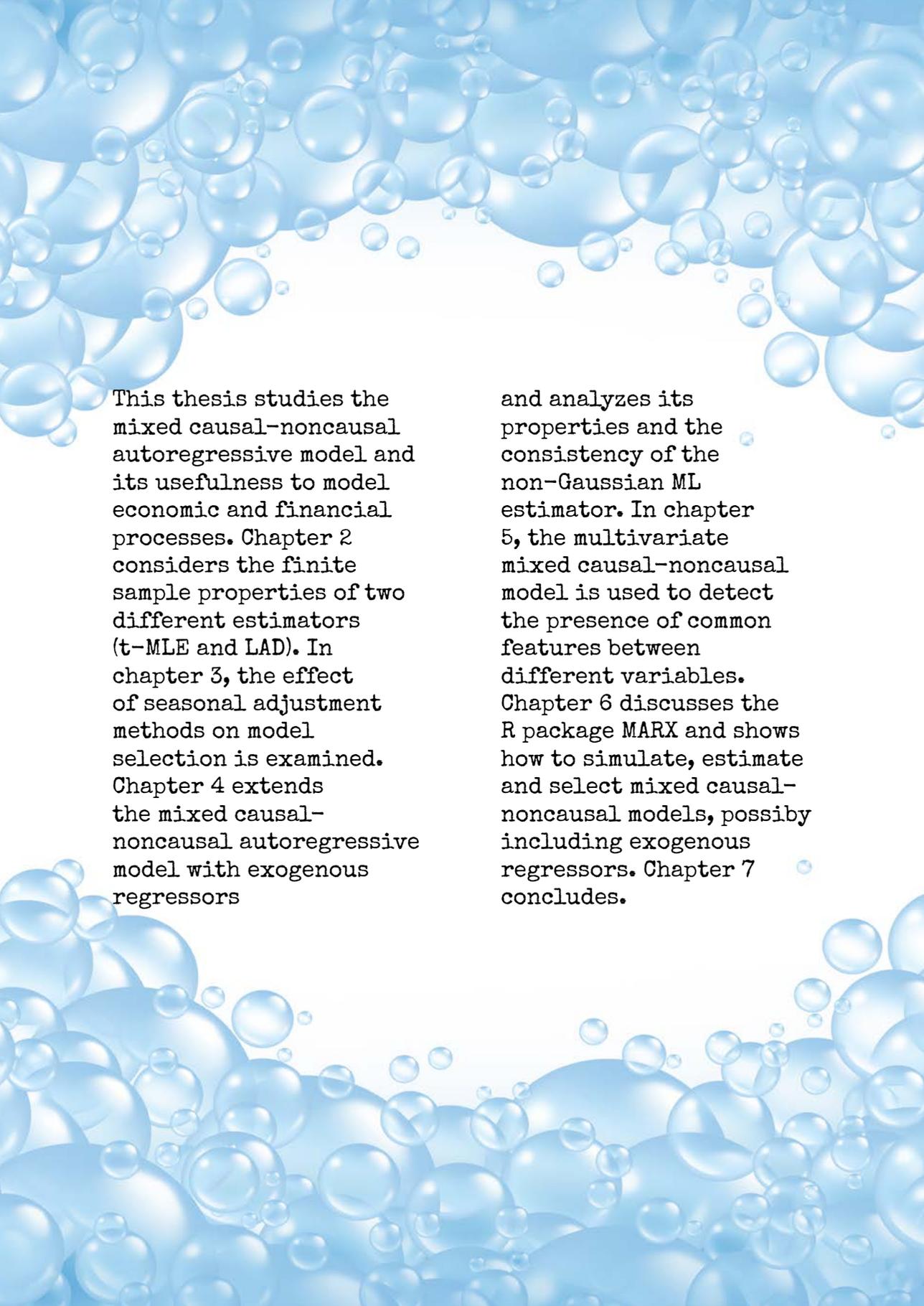
“Life isn’t about finding yourself. Life is about creating yourself.”

-George Bernard Shaw (1856-1950)

Johannes Maria Andreas (given name: Sean) Telg was born on 18 November, 1990 in Maastricht, The Netherlands. He attended high school between 2003 and 2009 at Bonnefanten College in Maastricht, where he received his Gymnasium diploma. Subsequently, he studied the bachelor Econometrics and Operations Research at Maastricht University. During his undergraduate studies, he spent one semester at Universidade Nova de Lisboa, Portugal as a visiting student. After completing his bachelor, Sean engaged in the two-year Economic and Financial Research Master - track Econometrics, for which he obtained his M.Sc. degree in August 2014.



After graduation, Sean joined the Department of Quantitative Economics as a Ph.D. candidate in September 2014, under the supervision of Prof. dr. Alain Hecq and Dr. Lenard Lieb. He spent three months at the Centre de Recherche en Economie et Statistique (CREST) in Paris as a visiting researcher. The results of his research are presented in this thesis. Sean presented his work at various international conferences, such as the Annual Symposium of the Society for Nonlinear Dynamics and Econometrics (SNDE) in Oslo and Paris, the Annual Netherlands Econometric Study Group Meeting (NESG) in Maastricht and Leuven, the International Conference on Computational and Financial Econometrics (CFE) in London and Sevilla, the summer school of the Society of Financial Econometric (SoFiE) in Brussels, the Meeting of the Econometric Society (ESEM) in Geneva and Lisbon and the Workshop on Advances in Quantitative Economics in Maastricht.



This thesis studies the mixed causal-noncausal autoregressive model and its usefulness to model economic and financial processes. Chapter 2 considers the finite sample properties of two different estimators (t-MLE and LAD). In chapter 3, the effect of seasonal adjustment methods on model selection is examined. Chapter 4 extends the mixed causal-noncausal autoregressive model with exogenous regressors

and analyzes its properties and the consistency of the non-Gaussian ML estimator. In chapter 5, the multivariate mixed causal-noncausal model is used to detect the presence of common features between different variables. Chapter 6 discusses the R package MARX and shows how to simulate, estimate and select mixed causal-noncausal models, possibly including exogenous regressors. Chapter 7 concludes.